OM of: US-09-494-297-2 to: N_Geneseq_0401:*   out_format : pfs

Date: Jun 7, 2001 12:15 AM

About: Results were produced by the GenCore software, version 4.5,
    Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn1_1/USPTO_spool/US09494297/runat_06062001_115735_12347/app_query.fasta_1.825
-DB=N_Geneseq_0401 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPEXT=4.000 -FGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09494297_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-494-297-2
Query length: 757
Database: N_Geneseq_0401:*
Database sequences: 678276
Database length: 291890651
Search time (sec): 149.980000

score_list:

Sequence	Std	Orig	ZScore	EScore	Len	!	Documentation
	Score						
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X12982	+	196.00	241.91	1.6e-05	15614	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:A30874	+	180.00	238.53	2.4e-05	2127	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X77593	+	178.00	223.16	0.0002	7650	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:V82021	+	176.00	230.41	6.8e-05	2718	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:V82021	+	161.00	204.61	0.0019	4249	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X20200	+	161.00	204.36	0.0019	4359	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X13179	+	161.00	194.10	0.0071	12445	!	
/cgn1_8/gcgdata/geneseqn/NA1997.DAT:T68844	+	154.00	192.00	0.0093	5547	!	
/cgn1_8/gcgdata/geneseqn/NA1998.DAT:V29930	+	152.00	189.14	0.0135	5547	!	
/cgn1_8/gcgdata/geneseqn/NA1997.DAT:T68843	+	152.00	186.13	0.0198	7551	!	
/cgn1_8/gcgdata/geneseqn/NA1998.DAT:V29928	+	152.00	186.13	0.0198	7551	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X77451	+	150.00	190.70	0.0110	3531	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:A97541	+	149.00	191.89	0.0095	2700	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X77593	+	149.00	188.18	0.0152	3945	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X20228	+	149.00	183.81	0.0267	6168	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X20230	+	149.00	183.81	0.0267	6168	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X20232	+	149.00	183.81	0.0267	6168	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X13101	+	149.00	169.25	0.1727	27360	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X13101	+	148.00	183.28	0.0719	43576	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:A68609	-	147.00	172.26	0.1175	15016	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X99560	+	146.00	182.37	0.0321	4612	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:Q24123	+	145.50	186.99	0.0178	2672	!	
/cgn1_8/gcgdata/geneseqn/NA1998.DAT:V39862	+	145.50	182.65	0.0310	3600	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:V04279	+	144.50	180.93	0.0386	2976	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:Z93534	+	142.00	180.93	0.0386	2976	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:Z51202	+	142.00	174.53	0.0878	3695	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:A68609	-	140.00	163.28	0.3719	27360	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:X20209	+	139.00	174.15	0.0921	3840	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:X20208	+	139.00	168.15	0.1991	7101	!	
/cgn1_8/gcgdata/geneseqn/NA1996.DAT:T29247	+	139.00	168.15	0.1991	7101	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X13104	+	139.00	158.51	0.6853	19031	!	
/cgn1_8/gcgdata/geneseqn/NA1998.DAT:V30561	+	138.00	158.00	0.6853	7642	-	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:V82026	-	137.50	155.85	0.9644	20072	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:Z06831	+	137.50	149.62	2.14	37948	!	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X13301	+	137.00	162.43	0.4143	9509	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:C42710	+	136.00	166.54	0.0679	1940	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:C50213	+	136.00	166.79	0.2369	5259	!	
/cgn1_8/gcgdata/geneseqn/NA2000.DAT:A55565	+	135.50	172.74	0.1105	2661	!	
/cgn1_8/gcgdata/geneseqn/NA1998.DAT:V52187	+	135.50	152.55	1.47	20986	-	
/cgn1_8/gcgdata/geneseqn/NA1997.DAT:T75487	+	135.00	168.84	0.1823	3687	-	
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X22748	+	135.00	168.84	0.1823	3687	-	

seq_name: /cgn1_8/gcgdata/geneseqn/NA1999.DAT:X12982

/cgn1_8/gcgdata/geneseqn/NA1999.DAT:A71797	+	135.00	168.84	3687
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X78246	+	135.00	168.62	0.1873	3768
/cgn1_8/gcgdata/geneseqn/NA1996.DAT:T08695	+	135.00	168.46	0.1913	3832
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X78245	+	135.00	165.88	0.2663	4988
/cgn1_8/gcgdata/geneseqn/NA1999.DAT:X99555	-	134.00	169.24	0.1731	3057

seq_documentation_block:
ID   X12982 standard; DNA; 15614 BP.
XX
AC   X12982;
XX
DT   19-MAR-1999   (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:45.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998; 98WO-US08985.
XX
PR   14-NOV-1997; 97US-0066009.
PR   06-MAY-1997; 97US-0044031.
PR   16-MAY-1997; 97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC, Dillon PJ, Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 419-427; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 15614 BP; 5535 A; 2567 C; 3169 G; 4326 T; 17 other;

alignment_scores:
    Quality: 196.00   Length: 706
    Ratio: 0.596   Gaps: 42
    Percent Similarity: 46.601   Percent Identity: 22.946

alignment_block:
US-09-494-297-2 x X12982

Align seg 1/1   to: X12982   from: 1   to: 15614

```
96    ValAsnLeuGluGlySer......ArgSerTyrGlnValTyrCysPheAs    110
      |||:||   |||:::
9070  GTGTCGCTCAACAGGCGCTACAATTAATAGTAATGAGATTTAT          9111

110   nLeuLysLysAlaPheProLeuGlySerAspSerSerValLysLysTrpI   127
      ::||||||||
9112  .......TTGGGTAAAGGGCGAAGAAAATTCAAATTCATT             9142

127   yrLysLysHisAspGlyIleSerThrLysPheGluAsp............   139
      ::||  ||||:||:|
9143  ATCAAGTACGT......ATTCAAACAGAGTCAGAAAACTTCAAACTTCAT    9186

140   .......TyrAlaMetSerProArgIleThrGlyAspGluLeuAsnGlnLy  154
             |||||:|
9187  TTTTGGTTATCAAATTGAATGGTCCGGACAACGTTTCAGCCCATTAGCCACCGGC  9236

154   sLeuArgAlaAlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleM  171
      |||:|  ::|
9237  CCCTGAAAAAGTTGATTTT......GGGGTTC                     9262

171   etGluGlyLeuGluProLeuAsnAlaIleArgValThrGlnGluAlaVal   187
      ::|||  |||
9263  CTTCGGAAAAGCACCT....GGCGTGAAGTTAAACGTGAAAAAATC       9306

188   TrpTyrTyrSerAspAsnAlaProIleSerGluAsnProAspGluSerPheLy  204
      |||  |:::|||||:|||||||:|
9307  TGGGAAGAGTATATGATCAAGACCCGACAAGCTCGGCACGATAATGTGATTTA  9356

204   sArgGluSerGluSerAsnLeuValSerThrSer..........         215
      |:||  ::|
9357  TGAAAATTAGTAGAAAAGCAAAGTAACTGACAGCCAACTGGGT          9406

216   .....GlnLeuSerLeuMetArgGlnValAlaLeuLeuGlnLeuIleAspPro  230
           ||||||
9407  ATATTAAATTATCAAAAACCAGAAAAATGATTACCCAGCAATAGTTGGGACGCGC  9456

231   AsnLeuAlaThrLysMetProLysGlnValProAspAspPheGln........  245
      |||  |||:|||::|
9457  AAAAAATGTAACCCAACTTTCCAAAACCCGGATGAAAGCTATCAAGAAGT    9506

246   .LeuSerIlePheGluSerGluSerGlyAspLysGlySerGlyTyrArgAsnLysGly  262
      |||:|||:::::|||||||||:||
9507  TCTTTGGGCTTCCCCAAAATACACAATCAAGGACAAAGCTTTCAAT....T  9550

262   yrGlnAsnLeuLeuSerGlyIleGlyLeuValProThrHisLysProProThrPro  278
      |||:|||  :::  |||
9551  ATCAAAACAACCCGTGAATTAGCA.....GTTCCTGGTTACAGTCAAGAAAAA  9597

279   GlyAspProProCysMetProProAsnGlnProGlnThrThrSerVal......  293
      ||||||
9598  ATCGACGGATACTACTTGGAAAAACACGAGCAGTTCAAGCCATTAGAATTT  9647

294   ...LeuIleArgGlyLysTyrAlaIleGlyIleAspTyrSerLysLeuLeuGluG  309
         :::||||::  ||
9648  AAAAGTAATCAAAAAATTCTTCTCCAGGTGAG......AAAAACTTAGTGG  9691

309   lyAlaThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGlnAlaArg   325
      ||||||  |:||:|
9692  GAGCCGTCTTTGAATTGACTGGTAAAAATGTT...........CAAACAACA  9732

326   ValPheSerSerAsnAsp............IleGlyGluArgIleGluGlu   338
      ::|  ||||
9733  TTAGTGGACAATTAAAGATGGTAGCTATTCCTTGCCAAAAGATGTGCGGCT  9782

338   uSerAspGly......ThrThrTyrThrLeuThrGluAsnSerProAlaGlyT  354
      |||  ||||||||||
9783  ACAAAAAAGGGGAACGCTATACAATTACACTGAAAACCAAGCCACCTGCAGGAC  9832

354   yrSerIleAlaGluProIleThrPheLysValGluAla........Gly    367
      ::|:|||||   |||
9833  ATGAGTTAGGCAAGAAAAACGACTTGGCAAATTGAGGTGGAGTGAGCAAGGC  9882

368   LysValTyrThrIleIleAspGlyLysValGlnIleGluAsnProAsnLysGl  384
      ||||||||:::|
9883  AAAGTA......AGCATCCGATGGACAAGAGAAGTGACCACCACAAATCAGT  9926

384   uIleValGluProTyrSerValGluAlaThrTyrAsnAspSerPheGluLuPheS  401
      ::  |||  |||  :::||::||::|
9927  TATT......CCATTGGAAATTGAA......AATAAAATTTTCTTGTTTGC  9964

401   erVal.............LeuThrThrGlnAsnTyrAlaLysPheTyrTyr   413
      :::  |||
9965  CAATCAGATTAGAAAAATAGAAAATACACCATTGCAAAAT............  9996

414   AlaLysAsnLysAsnGlySerSer.GlnValValValTyrCysPheAsnAlaA  430
      :|::   |||  |||  ::
9997  GGCAAACAAGTGAACATTAGCAGAGGCGACTTTTCGGTTCCAAAGAA      10042

430   spLeuLysSerProProAspSerGluAspGlyLyLysThrMetThrPro    446
         |||   :::
10043 .........AAAATGCTGCAAGGA                             10057

447   AspPheThrPhrGlyIleuValLysThrTyrThrHisIleIleAlaGlyArgAspIu  463
      ::::   |||
10058 AGTTACCAAAACTGTGGCAACTCAAAAAACAGATACTACAGGATTGCAGCTA  10107

463   uPheLysTyrThrValLysProArgAspThrAspAspThrPheLeuL      480
      |||
10108 TTTT......                                           10111

480   ysHisIleIleLysLysValIleLeuGlyLysGly...TyrArgGluLysGlyGln  495
      |||:::   |||   |||:|||
10112 ......AAATTAGTGAACCTGGTGAGTATCGA.                    10138

496   AlaIleLeuGluTyrSerGlyLeuThrGluThrGlnLeuuArgAlaAlaThrGl  512
      ::|||||   |||
10139 ATGGTGGAACAATCAGGA..........CCATTAGGCTACGACACTCT      10176

512   nLeuuAlaIleTyrTyrProPheThrAspSerAlaGluLuAspLysAspLysSL  529
      :::|||   |||
10177 TGCTGGAAATTATGAATTTACT......GTTGAAAA                 10207

529   euLysAspTyrHisGlyLysPheGlyAspMetAsnAspSerThrLeuAlaVal  545
      ::||
10208 .........TAIGGGAAA                                   10216

546   AlaLysIleLeuGluValGluTyrAlaLaGln.......AspSerAsnProPr  559
      ::|||||:||||||   |||   ||
10217 ......ATTCACTATGCAGGCAAAAAAATATTGAAGAAAAAAATGCCGCC    10254

559   oGlnLeuuThrAspLeuuAspPhePheIleProAsnAsnAsnLysTyrGlnGln  576
      :::   ||||
10255 AGAATGGACA                                           10264

576   erLeuuIleGlyThrGlnInrpHisProGluAspProLuuValAspIleIleArg  592
      ::||
10264 ......                                               10264

593   MetGluAspLysLysGluValIleProValThrHisIleAsnLeuThrLeuAr  609
      ::|||||||::
10265 ..........CTGACACATCAAAAATAATTTGAA                   10287

609   gLys..........ThrValThrGlyLeuAlaGlyTyrAspArgThrTyrAspP  623
      |:::
10288 ACCTTTTGACTTAACAGTTATTAAAAAACCGATAATACACG           10330

623   heHisPheGluIleLeuLysAsnAsnLysGlnInGluLeuuSerGln      639
      |||||||:::   ||
10331 .........CCACTTAAACGAGGCCAAAATTCCGTTTAACAGGA         10363

640   ThrValLysThrAspLysThrAsnLeuGluPhe...LysAspGlyLysVal   655
      |||||||:::|
10364 ......CCAGATACGGATAATTGAATTACCAAAAAGAGATGGCAAAGA      10401

655   aThr..........IleuAsnLeuuLysHisGlyGlu...SerLeuuT     666
      |||
10402 ......IleuAsnLeuuLysHisGlyGlu                        10413
```

```
10402 AACGGGATACTTTGTTTTGAAACTTAAAACCAGGGAAATATGTTCTAA 10451
  666 hrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysGluThrAsp 682
      ||||||||||||||:::|||  |||:   :::
10452 CAGAAACCTTACGCCAGAAGGATATCAGGGGTTAAAGAACCAATCGAA 10501
  683 ......SerGluGlyTyrLysValLysValAsnSerGlnGluValAl 696
            |||  |||||||||||||
10502 TTAATAATTCGTGAAGATGGTTCAGTCACGATAGATGGGGAAAAAGTAGC 10551
  696 aAsnAlaThrValSer.........LysThrGlyIleThrSerAspGluT 710
       :::||| ||||||:||| |
10552 AGATGTTTTAATTCGTGGAGAAGAATAAATCAAAATTACTTAGACGTTA 10601
  710 hrLeuAlaPheGluAsnAsnLysGluProValValProThrGlyValAsp 726
       ::||||  ||| |||||||||| |||||
10602 CG...AACCAAGCAAAGGTTCCTTTACCTGAAACTGGTGGCCATA 10642
  727 GlnLysIleAsnGlyTyrLeuAlaLeuIle.........ValIleAlaIg1 740
           |||||||    |||
10643 GGACGCTTGTGGTTTACTTGATAGCCATTAGTACACATTCGTGATAGCCGG 10692
  740 yIleSerLeuGlyIle 745
       |::||  |||
10693 TGTTTATACTCTTTATT 10708
```

seq_documentation_block:

seq_name: /cgn1_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A30874

```
ID   A30874 standard; DNA; 2127 BP.
XX
AC   A30874;
XX
DT   19-SEP-2000  (first entry)
XX
DE   Streptococcus equi fibronectin binding protein, FNZ, coding sequence.
XX
KW   Fibronectin binding protein; SFS; vaccine; horse; strangles; therapy;
KW   equine upper respiratory tract disease; S. equi infection; FNZ; ds.
XX
OS   Streptococcus equi.
XX
FH   Key            Location/Qualifiers
FT   CDS            108..1901
FT                  /*tag= a
FT                  /product- FNZ
XX
PN   WO2000037496-A1.
XX
PD   29-JUN-2000.
XX
PF   21-DEC-1999;   99WO-SE02448.
XX
PR   22-DEC-1998;   98SE-0004491.
XX
PI   Guss B,  Lindmark H,  Jacobsson K,  Frykberg L;
PA   (GUSS/) GUSS B.
PA   (LIND/) LINDMARK H.
PA   (JACO/) JACOBSSON K.
PA   (FRYK/) FRYKBERG L.
XX
PT   New protein useful for preparation of vaccines for treatment of
PT   strangles caused by Streptococcus equi infection, is able to bind to
PT   mammalian fibronectin -
XX
PS   Disclosure; Page 17a-17b; 34pp; English.
XX
CC   This sequence encodes the Streptococcus equi fibronectin binding
CC   protein FNZ. The FNZ sequence was used to isolate the S. equi fibronectin
CC   binding
```

```
CC   binding protein of the invention, designated SFS. SFS binds specifically
CC   to mammalian fibronectin or its analogues or fragments. The protein, its
CC   analogues or fragments may be used for the preparation of a vaccine that
CC   protects horses against strangles (a world-wide distributed and serious
CC   disease of the equine upper respiratory tract) caused by S. equi
CC   infection. The antibody and/or antiserum may also be used for the
CC   prophylactic or therapeutic treatment of S. equi infection in mammal,
CC   especially horses. The use of vaccines containing the fibronectin binding
CC   protein provides a more effective protection against S. equi infections,
CC   with fewer side effects.
XX
SQ   Sequence 2127 BP; 718 A; 415 C; 466 G; 528 T; 0 other;
```

alignment_scores:
```
  Quality:        180.00         Length:          716
  Ratio:          0.562          Gaps:             33
  Percent Similarity:  44.693    Percent Identity:  20.950
```

alignment_block:
US-09-494-297-2 x A30874

Align seg 1/1   to: A30874   from: 1 to: 2127
US-09-494-297-2 x A30874   ..

```
 68 SerGluTyrArgTrpTrpTyrGlyTyrGluSerTyrValArg...GlyHisPr  83
    ::::||| ::|||||||||||| ||||||:::||||| ||
201 GCAGAGCAGCTTTATTATGGGTGGTATGGATGGAGACGAGACAAAGTTCGCC 250

 83 oTyrTyrLysGlnPheArgValAlaHisAspLeuArgValAsnLeuGlu  100
    ||||:||| :::  :|||||  |||  ||
251 ATATTTT.....TTGTACGTATCGCCTAAAAATGCTCCAAAGCGTGAAT 294

100 lySerArgSerTyrGlnValTyrCysPheAsnLeuLysLysAlaPhePro 116
    ||||||:||||||||||||||||||||||: |||  ::|||||||
295 TAAAAGACGAGTAGTGTTGTTTATTGCTTTAACAAAAAATTGTATTGGCCA 344

117 LeuGlySerAspSerSerValLysLysTrp..............  126
     : :|||
345 GATCAATGGGAAATCTATATACGACAATTTAATGACATCAGATCTCCATA 394

127 ...............TyrLysLysHisAspGlyIleSerThrLysPheG 138
                  |||:::|||  :::  |||
395 TAACGAATTTACCTGTATATGAGAAAAAACTAGGATATGATGGTATATTTA 444

138 luAspTyrAlaMetSerProArgIleThrGlyAspGluLeuAsnGlnLys 154
    :::||||:||| |||  ::: |||
445 AACAATAATGCTCCAGATTACAAAAAAGAATATTAGTGATATTGCAAGTGCT 494

155 LeuArgAlaValMetThryrAsnGlyHisProGlnAsnAlaAsnGlyIleMe 171
    ||||  | ||::||||  :::  |||||:|||
495 TTGGTGGCAGTTTAAGTAATGGAATACCCCACTAACAAGTCACAACTATC 544

171 tGluGly.......LeuGluProLeuAsnAlaIleArgValThrGlnGluA 186
    |||||   :::||| |||||||||||| |
545 AACTAGCTACCATTAAAATAAATGATTCTTCTAGAAAAGTTTACTCAATTAG 594

186 laValTrpIleTyrTyrSerAspAsnAlaProIleSerAspAsnProAspGluSer 202
    |||:|||||||:||||||||||  ||| ||||
595 CCATTGGTTTPATTTTAGTGAT............AGTTTAACAAAAGAAATAC 632

203 PheLysArgGluSerGluSerAsnLeuValSerThrSerGlnLeuSerIle 219
    ||||  |||||||:::  ||||||
633 CTTAAAGATACAGGGTTGGTTATAACTTA............AACGATAT 667

219 uMetArgGlnAlaLeuLysGlnLeuGluAspProAsnLeuAlaThrLysM 236
    ||||||:||||  |||||:::||:  ::|||:
668 GGAAAAAAAGGCTTTAGATTTTTTTAATCAGTAAAGGAGGAGGATCTAAGC 717

236 etProLysGlnValProAspAspPheGlnLeuSerIlePheGluSerGlu 252
    ||  |||:::||||||||:|||::|
718 TT...AAATCAGAGCAGAGTAATTTACTCATTGATATATTTTTATGTTATCAA 764

253 AspLysGlyAspLysTyrAsnLysGlyTyrGlnAsnLeuLeuSerGlyGl 269
    ::|||||||||||  |||||:::  ||||:||::|
```

```
              :::                           |||||||||||||||||||:::::
765  AGTGGCGGGCGGCCATATGAAAGATTACCAAAAAAGTTCTCGGCTCTAC                                   814
       |||
269  yLeuValProThrLysProProThrProGlyAspProProMetProProA                                  286
     ||:||:|||
815  CTTAATTCCTAAAGAACCGCTAAAGCCT                                                         842
              |||
286  snGlnProGlnThrThrSerValLeuIleArgLysTyrAlaIleGlyAsp                                   302
                                   :::|||
843  .........                                                  CAGCTAGGTGGT               854
              |||
303  TyrSerLysLeuLeuGluGlyValGlyAlaThrLeuGlnLeuThrGlyAspAsnVa                             319
       |||
855  TTTAGT                                        GGACATAATGG                            871
              |||
319  lAsnSerPheGlnAlaArgValPheSerSerAsnAspPheIleGlyGluArgI                                336
     ||:||:||||::|::                             |||
872  AAATGGATTAAGCGGCCCTTGAAGGAGGATCATCAGGTTCACAAGAAACTA                                  921
              |||||:|
336  leGluLeuSerAspGlyThrTyrThrThrLeuThrGluLeuAsnSerProAla                                352
                                       :::|||
922  ATGAA......                    GATGTGAAGAAAAGGACTT.....  ATAGGTTTCCAGGA               959
              |||
353  GlyTyrSerIleAlaGlu..........        ProIleThr..  PheLysVa                             364
     ||||
960  GGACTCTCAGGAGCGAGGGCAAACGAGATCCTTTGCCAGGATTGAAGG                                     1009
              |||||:|:||
364  lGluAlaGlyLysValTyrThrIleIleGluAspGlyLysGlnIleGluAsnP                                381
     |:|||||||||
1010 TGAGGGCTGGTGCACCTGGATACA.......                  CCTCAAAAGC                          1041
              |||:
381  roAsnLysGlyIleValGluProThrSerValGluAlaLeuTyrArgAsnAspPhe                             397
                                       |||
1042 CTAATGATTCCATTG........            CAAGGTCTTGAAGGCGGTAAC.                             1076
              :::
398  GluGluProSerValLeuThrThrGlnAsnTyrArgAlaLysPheTyrTyrTyrAl                              414
       |||                                          :::
1077 .........        TCTCCTAAGTAGTAGAACAAAAACTAAGGTAGT.........   AC                      1108
              |||
414  aLysAsnLysAspGlySerSerGlnValValValTyrCysPheAsnAlaAspL                                431
     |:|||:||:                  ||:||:|||
1109 CGAAGGATATCATGGTCAAATCAGGCATTCTTGAG......   GAAACCGAAG                                1152
              |||
431  euLysSerProGluAspSerGluLlsAspGlyLysLysThrMetThrProAsp                                447
     ::||||||||
1153 ATACTAACCACCTGGTATCATACTAGGCGGGTCTCAGGAAAATGTCGAA..                                  1199
              |||
448  PheThrHrThrGlyGluValLysTyrThrHisIleAlaGlyv..........                                 460
                                       :::|||
1200 .    ACGCATGAAGATACTAGAAAACCCTCATCTGATGGGATCGGCGCGCG                                  1246
              |||
461  ......  ArgAspLeuPheLysTyrThrHrValLysProArg..........                                471
                                   :::
1247 TCTAGCTGGCGAATCAGGAGAACGACACCTAAACCAGGACAAACCGGCG                                    1296
              |||
472  ..........              AspThrAspProAspThrPheLeuLysHisIle                             482
                                       ||||||
1297 GGCAAGGACCACTCATCGAGCAACACAGGGATACA.                                                 1331
              |||
483  LysLysValIleGluGlyLysGlyTyrArgGlyLysGlyGlnAlaAlaIleGluThrY                           499
                                       |||||||||
1332 .........                CAAAAGGCARGTCTGGACAATC                                       1354
              |:||
499  rSerGlyLeuThrGlnThrGlnLeuArgGlnAlaLaAThrGlnLeuAlaIleT                                516
     ::||                |||
1355 TGGTGGCCACTATCGAGTCAAGAAACACCAAAAAGCCGGAGGTCATGATTG                                  1404
              |||
516  yrTyrPheThrArgAspSerArgAlaGluLeuAspLysAspLysSerLeuLysAspTyr                           532
                                   :::|||
1405 GTGGTCAGGAGGACAAAACCATCGAGACAACAGAGGACGACCACAAAAA                                    1448
              |||
533  HisGlyPheGlyAspMetAsnAspSerThrLeuAlaValAlaLysIleLe                                   549
     ::|||:::                                   ::|||::::
```

```
              :::                           GGCATGTCTCGGACAATCTGCGGTACTATCGAG             1481
1449 .....    GGCATGTCTCGGACAATCTGCGGTACTATCGAG.........                                  1481
              |||
549  uValGluThrAlaGlnAspSerAsnProGroGlnLeuThrAspLeuAspP                                   566
     |||::||:||
1482 ..........              TCAGAGGACACATAAGAAAACCTGAGGTC.........                        1508
              |||
566  hePheIleProAsnAsnLysTyrGlnSerLeuIleGlyThrGlnTrp                                      582
     ::|||||||||                     |||
1509 ..........              ....    ATGATTGGGTGGTCAG...                                   1523
              |||
583  HisProGluAspLeuValAspIleIleLeuArgMetGluAspLysLysGluvA                                599
                                       :::||||
1524 ..........              .GGACAAAT                                                     1531
              |||
599  IlleProValThrHisIleAsnLeuThrLeuArgLysThrValThrGlyLeuA                                616
     |||:||:|||                     |||::::
1532 CATCGGACTTCTTCTCGAAAC...    ACCCAATCAGGTAATGTCTGGGCAGT                                1575
              ||||
616  laGlyAspArgThrLysAspPheHisPheGluIleGluLeuLysAsnAsn                                   632
     ::||||||:|                     :::||||||
1576 CTGGTGACACTACG......            GTAATTGAG.........                                   1598
              |||
633  LysGlnGluLeuLeuSerGlnThrValLysThrSerThrAsnLeuGl                                      649
     |||::|||||:::|:|::                          :::|||
1599 uPheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGluSerLeuR                                   1621
     ::::                           GATACCAAGAAGTCTGAGAATAAT                              
1622 CATTGGTGGGCAGGACAAATCAATCGACTTCTCTGAGGATAACTCAGCCCG                                  1671
              |||
649  rValSerLysThrGlyIleThrSerAspGluThrLeuAlaPheGluGluAsnA                                666
     ::|||::|||
666  hrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysGluThrHisAsp                                682
     :::|||::|||                     |||
1672 GTAWGTCTGGT......            CAAATCTGGAGGCACTACAATTGTCGAAGACACCAAG                     1718
              |||
683  SerGluGlyThrTyrLysValLysValAlaAsnSerGlnLuValAlaAsnAlaTh                              699
     ::|||:|||                       :::::|||
699  rValSerLysThrGlyIleThrSerAspGluThrLeuAlaPheGluGluAsnA                                716
                                   GACGA.......  AAACCTA                                  1776
1719 AAGCCGACAACCTAAGCCTAAAACCTGCCACCTGCGCCAAATTGTTAAT.......                             1763
              |||
716  snLysGluProValValProThrGlyValAlaAspGlnLys........                                   728
     |||||||                         :::|||
1777 ACAAAGGCACTCAICTCCCACAGACAAGTGAATATGAAGACAACTCACCCTA                                 1826
              |||
729  ....    IleAsnGlyTyrLeuuAlaLeuIleValIleuAlaGlyIleSerLeu                              743
     |||:||                          :::::|||::::::|||
1827 AGCATCATCGGTCGCAATGTCAAATGCCTGCTTGTCCTTATGTCTGTCTCTA                                 1874
```

```
FT   RBS             127..136
FT                   /*tag= c
FT   CDS             141..2837
FT                   /*tag= d
FT                   /gene= lbpB
FT                   /product= LBP2
FT                   /transl_except= (pos:2031..2032, aa:Asn)
FT                   /note= "this codon has an apparent 1 codon deletion
FT                   which alters the reading frame"
FT                   /transl_except= (pos:2036..2039, aa:Ala)
FT                   /note= "this codon has an apparent 1 codon
FT                   insertion which alters the reading frame"
FT                   /note= "lbpB is specifically claimed in Claim 8;
FT                   encodes W89413"
FT   misc_feature    632
FT                   /*tag= e
FT                   /note= "this base is not present in the sequence
FT                   given in the specification, the nucleotide
FT                   has been included to maintain the reading
FT                   frame as part of a codon (AAR) for Lys"
FT   -35_signal      2973..2978
FT                   /*tag= f
FT   -10_signal      2990..2995
FT                   /*tag= g
FT   RBS             3006..3015
FT                   /*tag= h
FT   CDS             3021..6023
FT                   /*tag= i
FT                   /gene= lbpA
FT                   /product= LBP1
FT                   /note= "specifically claimed in Claim 8; encodes
FT                   W89414"
FT   CDS             3066..6023
FT                   /*tag= j
FT                   /gene= lbpA
FT                   /product= LBP1
FT                   /note= "specifically claimed in claim 8; encodes
FT                   W89415"
FT   CDS             6025..7650
FT                   /*tag= k
FT                   /gene= orf3
FT                   /note= "specifically claimed in Claim 7; encodes
FT                   W89416"
```

```
CC   W89416).  The lfr locus was identified following generation of a M.
CC   catarrhalis strain 4223 genomic DNA library and screening with
CC   specific hybridisation probes.  The genes and DNA sequences of the
CC   lfr locus are useful for diagnosis, immunisation, and the
CC   generation of diagnostic and immunological reagents.  Immunogenic
CC   compositions, including vaccines, based upon expressed recombinant
CC   Lbp1 and/or Lbp2 and/or ORF3, portions of these or their analogues,
CC   can be prepared for prevention of diseases caused by Moraxella. M.
CC   catarrhalis is a causative agent of otitis media and has been
CC   associated with sinusitis, conjuctivitis and inflammatory diseases
CC   of the lower respiratory tract, such as pneumonia, chronic
CC   bronchitis, tracheitis and emphysema.
XX
SQ   Sequence 7650 BP; 2410 A; 1741 C; 1636 G; 1862 T; 1 other;
XX
alignment_scores:
     Quality:          178.00
     Ratio:            0.467
Percent Similarity:    41.503     Percent Identity:  18.627
alignment_block:
US-09-494-297-2 x V82019

Align seg 1/1   to: V82019    from: 1   to: 7650

  4  ThrArgPheProAsnLysLeuValAsnThrLeuAsnThrGlnArgValLeuSe   20
     |||||  ||||      ||||||      :::    |||||||||||
375  ACACTGACCCAAAGACCGCGATAAACAACAACCACTGACACAA            413

 20  rLysAsnSerLysArgPheThrValThrLeuValGlyValPheLeuMetI     37
       |||:|||  ||      ||||:  :::    |   ||    ||:|||
414  ..GCACAAAAGACCGCCGCTGCCGCGCAGGGTTTTTTGTGATG.          452

 37  lePheAlaLeuValThrSerMetValGlyAlaLysThrValPheGlyLeu     53
                                                    |||
453                                              GGTAAA     458

 53  ...                                                    69
459  ATTCGTGATACCAGCGCCAAAAAAT...GACCCAGATTATAGCCAATGATTT  505

 70  ....TyrArgTrpTyrGly......TyrGluSerTyrValArgGlyHisP     83
       :::|    |||     |||       :::               :||
506  AGTACAGCAGTGGCAAGGCAAATTATATATGTTGGGTATTGATGCCCAATCGCC 555

 83  ro..                                                   83
     ||
 83  ro..                                                   95
556  CAGATGGCATGGCACAGGTAAAAACTTGCGTCAGCCCATCACCGCCAAT      605

 84  ......TyrTyrLysGlnPheArgValAlaHisAspLeuAr              95
            ||:::|||                  :::||
                  ...TyrTyrLysGlnPheArgValAlaHisAspLeuAr... 655

 95  gValAsnLeuGluGlyGlySerArgSerTyrGlnValTyrCysPheAsnLeuL  112
     :||:||||                   :::||||||||
606  GACATCAAACCCTTGTATTTTAACAARTTCCCTGCATTGTCGTCGATTGCA    655

112  ysLysAlaPheProLeuGlySerAspSerSerValLysLys......Trp    126
     |||:::|                              ::::||
656  TTTAGACAGTGAACCGCCACCGT...TTTGACCCCAAAAAGCTAAACACCA    702

127  TyrTyrLysHisAspGlyIleSerThrLysPheGluAspTyrAlaMetSe    143
     |||  ||||||  |||                     ||||||
703  TTAAAGTGTATGGGTTATGGCAACTTAACACACACCCTCTAAAAACAACACT   752

143  rPro...............................ArgIleThrGlyAspGluL 151
     :||:                               ||:::||||
753  TACATCAATCATCACCAGCGTGATAATATAAGAGAAAAATAAACAAGGCCGTTGCA 802

151  euAsnGlnLys....................................        154
     ||::||:||||
803  CCCTTATGAAAATATCCGTTTTGGGTATCTTGAACTACAAGGAAGCAGTC    852

853  TGACCCAAAAAAAAATGCGGATACTCCAAATGACAAAGACCGCCATTCCAAA  902
```

```
WO9855506-A2.
XX
PN   WO9855506-A2.
PD   10-DEC-1998.
PF   02-JUN-1998;  98WO-CA00544.
PR   08-MAY-1998;  98US-0074658.
PR   03-JUN-1997;  97US-0867941.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
PI   Du R,  Klein MH,  Loosmore SM,  Wang Q,  Yang Y;
XX
DR   WPI; 1999-070266/06.
DR   P-PSDB; W89413, W89414, W89415, W89416.
XX
PT   Lactoferrin receptor genes from Moraxella, especially M. catarrhalis
PT   - useful to diagnose Moraxella infection e.g. to detect otitis media
PT   due to M. catarrhalis infection and to immunise against such
PT   infections
XX
PS   Claim 8; Fig 2A-2P'; 202pp; English.
XX
CC   This polynucleotide comprises the lactoferrin receptor (lfr) locus
CC   of Moraxella catarrhalis (Branhamella catarrhalis) 4223. There are
CC   3 tandem genes in locus, identified as lbpA, lbpB (alternative
CC   start codons) and orf3A, respectively encoding lactoferrin binding
CC   protein 2 (Lbp2, see W89414 and W89415) and open reading frame
CC   protein 3 (ORF3, see
```

```
155  ...LeuArgAlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIl 170
          :::||||||||||||||||:::::
903  CCCAtGCCCAttTGTTTtaTCACGGA......GAAAACCCCAGCAGCCA 946

170  eMetGlyIleGluProLeuAsn.......                     178
          :::
947  GCTGCCCAGTGCTGGTAAAtTTAACTACACAGGCAACTGGCTGTACCTAA 996

179  .......AlaIleArgValThrGlnGluAlaValTrp             188
                  ||||:::       ||||
997  GTGAtGTCAAAAACCCCTGCCACTTTCAGCAtCAGAtGAtGAGTGGGG   1046

189  TyrTyrSerAspAsnAlaProIleSerAsnProAspGluSerPheLysAr 205
          |||||
1047 GTCTAtCTCAAtGCCAGTGGCAAAtCCAAT                     1076

205  gGluSerGluSerAsnLeuValSerThrSerGlnLeuSerLeu        219
          |||:::|||||||||:::|||||:::
1077 GAGGGCGAtGTCGTTCAGTCGCCGCCCACAtTTACTCAAACGGCT      1119

219  .......                                            219
1120 TTCAAtATAAGCACAGCCTGCCACTTATCAGGTGGGAtTTGACACAAAC  1169

220  .MetArgGlnAlaLeuLysGlnLeuIleLeuAspProAsnLeuAlaThrLy 235
          |||:::|||||||
1170 TCAtTAACAGGCAAGCTGTCTTATTATGCAAAtCCAACAACTGC       1219

235  sMetProLysGlnValProAspAspPhe........G              245
1220 CCAAGGCAAAtACATCAAAAAGCCAAtTTGACTACTACCACAAAGTCAAtG 1269

245  InLeuSerIlePheGluSerGluAspLys.......GlyAspLysTyrAsn 259
          |||:::
1270 AAACCGAtGTGTAtCAAAAtTGACCAAAAAtCAAACGGCAACGGCTTCGTC 1319

260  LysGlyThrTyrGlnAsnLeuLeuSerGlyGlyLeu...            270
1320 GGTACCGCCAAAtCTTTGGtTAAtTGAGAACACAGAAAACCGCACCTTTTAT 1369

271  .......ValProThrLysProProThrProGlyAspProMetPro      285
          :::                                      |
1370 CAAAGAGACTGTTCTCCAAAAAAGCCAAtCCCAAtATACCCA......AACC 1413

285  roAsnGlnProGlnThrThrSerValLeuIleLeuArgLysTyrAlaIleGly 301
          ||||||
1414 CTAAT                                              1418

302  AspThrTyrSerLysLeuGlyLeuGlyLysLeuTyr.laThrLeuGlnIleuThrGlyAspAs 318
          ||||||||||||
1419 .......TCAGACACGCTGTAtTTAGTGCAGGCTTTGTTTAtGGTGAGTCGGGCGAtGA 1462

318  nValAsnSerPheGlnAlaAlaArgValPheSerSerAsnAsp.......  331
          ||||||||
1463 GCTG                                               1500

332  .......IleGlyLeuArgIleGluLeuSerAspGlyThrThrLeu      345
          ::||||  :::||||
1501 AtGTGGTGTCTTTGGtGGtTAAACGAGACAAAACCTGTCGCCACA       1550

346  ThrGluLeuAsnSerProAlaGlyThrTyrSerIleAlaGluProIleThrPh 362
          ||||||
1551 AAAACGGTGTAtTTTAGtGCAGGCTTTGAA                     1594

362  eLysValGlnAlaGlyLysValThrTyrThrIleIleAspSerGlyAsnIleG 379
          ||||||:::|||
1595 TTTTGTGGAtAAtGAAACGAtTGGCAGAAtTAAtTAACAGCAAAAAGTTAA 1644

379  luAsnProAsnLysGlnIleValGlu.......ProTyrSerVal       391
          :::|||       ||||
1645 AtGAtGGCTGAtGAAtGAGAAAAAtTGAtTAAtGGTGtATATTCCTACCAGTGAt 1694
```

```
392  GluAlaTyrArgAsnAspPhe.......GluGluPheSerVa          402
          |||       |||||||:::
1695 GAACGCGtATGAtGAGAATTTCCTTGGGGGCGGCGAAAAAAAGCAGAATTCACCAA 1744

402  lLeuThrThrGlnAsn.......TyrA                         409
          :::
1745 AAAAGTCAGCACGACGACCCCAAGCCGTGCCCACCTTAtTTTGGGCACATG 1794

409  laLysPheThrTyrArgAlaLysAsnLysAsnGlySerSerGlnValValTyr 425
          ||||||:::       :::
1795 ATAAAtTTTAtTTTT.......AAtGGCAACTATTATGACCTAtCA       1832

426  CysPheAsnAlaAspLeuLysSerProProAspSerGluAspGlyGlyLyLy 442
          ||||||:::
1833 GCCAGCAGTGTTGATAAAtTGGCCCCTGCCCGAtGCGTCTCAAAGCCAACCA 1882

442  sThrMetThrProAspPheThrThrArgGlyGluVal               453
          :::::
1883 AtCCATTAAAGAAAAAtACCCTAAtGCCACACTAAAtTAAGGACAACCAAG 1932

453  .......                                             453
1933 TTACCGCCAtCGTGCTACAAGAGAGCCAAAGAtAAtAAGCCTTAtACCGCC 1982

454  .......LysTyrThrHisIleAla...GlyArgAspLeuPhe..       464
          :::|||  |||:::||:::
1983 AtTCGTGCCAAAAGCTAtCACGCACATCAGTTTTGGCGAGACCGTGTATAC 2032

465  .......LysTyrThrValLysProArgAspT                    473
          :::|||
2033 GAtGCCAAAACCAAAACCCCAACACACGCAGTTAtTTTTGCAAGGCGGGTAGGG 2082

473  hrAspProAspThrPheLeuLysHisIleLeuLysLysValIle...     486
          |||
2083 CAGAtACCAGCACCACCACGCTGCCCAAGGCAGGTAAAtTCACTTACACGGGT 2132

487  .......GluLysGlyThrArgArgGluLy                      493
          :::|||||||||
2133 CTTTGGGCAGGCGTAtCTCTTATCCAAAAAAAGGACAAAGGTTAtAGCAAtTA 2182

493  sGlyGlnAlaIleGluLysTyrSerGly...                     505
          :::|||||||||
2183 TGAAGAAACCAtCAAGAAAAAGGCCAtCAAGAAtTATCtGTTAACCGGAAG   2232

505  hrGlnLeuArgAlaAlaThrGlnLeuLeuAlaIleetyrTyrTyrPheThr 519
          |||:::
2233 AC.......                                           2246

520  .......AspSerAlaGluLeuAspLysAspLysSerLeuLysAspPY    532
          |||  :::
2247 GAtGAtGACGAtGACAtTTGACCGCCAtTCtGAtTCACAAGAtGAtGAtGC   2296

532  rHisGlyPheGlyAspMetAsnAspSerThrLeuAlaValAlaAlaLysIle 549
          ||||||
2297 ACAtGGCGGAtGAtGAtTTG.......AttG                     2319

549  euValGluTyrAlaGlnAspSerAsnProGlnLeuThrAspLeuAsp      565
          :::|||||||||||:::||||
2320 CAtCtGAtGAtTCACAAGAtGAtGCGCAGAAtGGCAGtACGACGAtTCAGAt 2369

566  PhePheIleProAsnAsnAsnLysTyrGlnSerLeuIleGlyThrGlnTr   582
          |||  :::||||
2370 GAtTTG.......GGTGAtGGTGCAGAtGACGACGCCCCGCAGGCAAAGTGtA 2413

582  pHisProGluAspLeuValAspIleIleArgMetGluAspProLysLysGluV 599
          :|||
2414 TCAtGCAGGTAAt.......AtTCGCCCtGAATTtGAAAACAAAt        2451

599  alIleProVal.......ThrHisAsnLeuThrLeu...             608
          |||  |||:::
2452 ACTTCCCCATTAAtGAGCCTACtCAtGAAAAAACCTTTGCCCtAGAtGGT   2501

609  .......ArgLysThrValThrGl                            614
```

```
seq_name: /cgn1_8/gcgdata/geneseq/NA1999.DAT:V82021

seq_documentation_block:
ID   V82021 standard; DNA; 2718 BP.
XX
AC   V82021;
XX
DT   21-JUN-1999 (first entry)
XX
DE   Moraxella catarrhalis VH19 lbpB gene.
XX
KW   Lactoferrin receptor; lactoferrin binding protein; LBP1; LBP2;
KW   lbpA gene; lbpB gene; ORF3; infection; otitis media; sinusitis;
KW   conjunctivitis; pneumonia; bronchitis; tracheitis; emphysema;
KW   diagnosis; therapy; vaccine; Branhamella catarrhalis; ss.
XX
OS   Moraxella catarrhalis.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..2718
FT                   /*tag= a
FT                   /transl_except= (pos:844..846, aa:Thr)
FT                   /transl_except= (pos:847..849, aa:Thr)
FT                   /transl_except= (pos:850..852, aa:Thr)
FT                   /transl_except= (pos:853..855, aa:Thr)
FT                   /transl_except= (pos:856..858, aa:Thr)
FT                   /transl_except= (pos:859..861, aa:Thr)
FT                   /transl_except= (pos:862..864, aa:Thr)
FT                   /transl_except= (pos:865..867, aa:Thr)
FT                   /transl_except= (pos:868..870, aa:Thr)
FT                   /transl_except= (pos:871..873, aa:Thr)
FT                   /transl_except= (pos:874..876, aa:Thr)
FT                   /transl_except= (pos:877..879, aa:Thr)
FT                   /transl_except= (pos:880..882, aa:Thr)
FT                   /transl_except= (pos:886..888, aa:Thr)
FT                   /transl_except= (pos:889..891, aa:Thr)
FT                   /transl_except= (pos:892..894, aa:Thr)
FT                   /transl_except= (pos:895..897, aa:Thr)
```

```
PT   Lactoferrin receptor genes from Moraxella, especially M. catarrhalis
PT   - useful to diagnose Moraxella infection e.g. to detect otitis media
PT   due to M. catarrhalis infection and to immunise against such
PT   infections
XX
PS   Claim 8; Fig 16A-O; 202pp; English.
XX
CC   This polynucleotide comprises the lactoferrin binding protein lbpB
CC   gene of Moraxella catarrhalis (Branhamella catarrhalis) strain VH19.
CC   It encodes lactoferrin binding protein 2 (Lbp2, see W89421). The
CC   genes and DNA sequences of the lactoferrin receptor (lfr) locus of
CC   Moraxella are useful for diagnosis, immunisation, and the
CC   generation of diagnostic and immunological reagents. Immunogenic
CC   compositions, including vaccines, based upon expressed recombinant
CC   Lbp1 and/or Lbp2 and/or ORF3, portions of these, or their
CC   analogues, can be prepared for prevention of diseases caused by
CC   Moraxella. M. catarrhalis is a causative agent of otitis media and
CC   has been associated with sinusitis, conjunctivitis and inflammatory
CC   diseases of the lower respiratory tract, such as pneumonia, chronic
CC   bronchitis, tracheitis and emphysema.
XX
SQ   Sequence 2718 BP; 958 A; 585 C; 547 G; 628 T; 0 other;
```

alignment_scores:
```
                  Quality:   176.00          Length:    796
                  Ratio:     0.515           Gaps:       40
Percent Similarity: 42.965         Percent Identity: 20.603
```

alignment_block:
US-09-494-297-2 x V82021

```
Align seq 1/1   to: V82021    from: 1   to: 2718

        84 TyrTyrLysGlnPheArgValAlaHisAspLeuArgValAsnLeuGluGl 100
           |||:::||||   ||:||                    :::||||||:|||
       493 TATTTTGATAAATTCCCAAAATATCCGATCTG......CACCTAGAAAA 536

       100 ySerArgSerTyrgInValTyrCysPheAsnLeuLysLysAla....... 114
           ||:|:|:::|                |  ||||||||||
       537 CAGCCAG.......CATGTGTTTGATGCTAAAAAAGCAAATAACA      574

       115 ........PheProLeuGlySerAspSerSerValLysLys......... 125
           :::|||:|  ||||:::                       ||||||||
       575 TCAAAATATATGGTTATGGTGCATTGTCATCACCCTGCCAAAAACCCAACC 624

       126 ........TrpTyrLysHisAspGlyIleSerThrLys.....PheGl 138
           ||:|:||||||:||:|||||||||||||   ||:::|      :|
       625 TACATGAATTATCAACAAGAACAAAAACATCAAAAACAAAACCAGGCGA 674

       138 uAspTyrAla.........MetSerProArgIleThr.........  147
           :|||||||               ||  |
       675 TGATTATCAAAACAACATTCGTTTTGGCTATAGGACCTAAGAGAGTGGACC 724

       148 ..........GlyAspGluLeuAsnGlnLysLeuArgAlaValMet  159
                   : :||  |||  |  :::|||::::||:|
       725 TAAATAAAAAAGGTGCCAGACACCCCAGACGACAACCGTGCCATCATT 774

       160 TyrAsnGly.........HisProGlnAsnAlaAsnGlyIleMe  171
           :::|:               |||  |  |:||||||  :  ::
       775 TTCACCACACCTACTTATTTATTATGGTGAGAGAATGCCAGCACCCATCT 824
```

```
2502 AAAAATAAAGCTAAGTTTGATGTGGATTTTGACACCAACAGCCTAACTGG
          ::::::::||||
       :::::::||||
2551

2552 TAAATTAAACGATGAGAGAGGTGATATCGTCTTTGATATC......AAA 2595
         |||:::||| ::
2595

2596 ATGGCAAAATTGATGGCACAGGCGTTACCGCCAAAGCCGATGTGCCAAAC 2645

2646 TATCGTGAAGAAGTGGGT..............ACAACAAGGTGGC 2678

2679 ..................GGTTTCTTATACAACATCAAAGATA 2703

2704 TTGATGTCAAGGGCAATTTTTTGGCACAAATGCCGAGAGTTGGCAGGG 2753

2754 CAGTTACGTACGACACAAAGGCCGATGGGCATCAATGACACCGCGAAAAAGC 2803

2804 AGGG 2807
```

```
2502 AAAAAT...  :::::::::||||||
614  yLeuAlaGlyAspArgThrLysAspPheHisPheGluIleGluLysA 631
2552 TAAATTAAACGATGAGAGAGGTGATATCGTCTTTGATATC......AAA 2595
631  snAsnLysGlnGluLeuSerGlnThrValLysThrAspLysThrAsn 647
           |:::::||| ::
2596 ATGGCAAAATTGATGGCACAGGCGTTACCGCCAAAGCCGATGTGCCAAAC 2645
648  LeuGluPheLysAspGlyLysAlaThrIleAsnLeuHisGlyGluSe 664
2646 TATCGTGAAGAAGTGGGT..............ACAACAAGGTGGC 2678
664  rLeuThrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysGluT 681
2679 ..................GGTTTCTTATACAACATCAAAGATA 2703
681  hrAspSerGluGlyTyrLysValLysValAsnSerGlnGluValAla... 696
2704 TTGATGTCAAGGGCAATTTTTTGGCACAAATGCCGAGAGTTGGCAGGG 2753
697  ..................AsnAlaThrValSerLysTh 703
2754 CAGTTACGTACGACACAAAGGCCGATGGGCATCAATGACACCGCGAAAAAGC 2803
703  rGly 704
2804 AGGG 2807
```

```
 171 tGluGlyGluProLeuAsnAlaIleArgValThrGlnGluAlaValT 188
     ::: |||
 825 GCCA.......AAGGCGGTAAAATTGACTATGAGGGCAATT 859

 188 rpTyrTyrSerAspAsnAlaPro...........IleSerAsnProAsp 200
     |||                             |||
 860 GGTTGTATCTGACGAGTCAAAAAACGCCCANTTTTAGAATAAAACAGAC 909

 201 Glu...........SerPheLysArgGluSerGluSerAs 210
     :::                ::::::::::::::
 910 GATAAAGTAGGCGACTTATTTTAACTCAACCAGAAAATCAAATGAAGGCGA 959

 210 nLeuValSerThrSerGlnLeuSerLeu.............. 219
     ||||||||||||:::::::|| |||
 960 TTTGGTGCAGTGCAGCACACATTTATCTAAACAGCTTTAAATATAAACACA 1009

 220 ....................MetArgGlnAla 223
                         :::::
1010 CCCCGGCGCCACTTATAGGGTGGACTTTGATCAAATATACCCTAAAAGGCAA 1059

 224 LeuLysGlnLeuIleAspProAsnLeuAlaThrLysMetProLysGlnVa 240
     |::|||        ||           :::||
1060 TTGTCTTTATTATNGACAACCCAAACAAGCAAACCGATGGGGCGTTATAT 1109

 240 lProAspAspPhe...............GlnLeuSerIlePheG 250
     ::::||||                    :::::::::|||
1110 CAGAAGTCAGTTTGATACCGACAAAAAAGGTCAATGAAGCCGANGTCTATG 1159

 250 luSerGluAspLys.......GlyAspLysTyrAsnLysGlyTyrGlnAsn 264
     |:::  |||             :::::::::::::
1160 AGATTGACGCCAAGATTAATGGCAAACCGGCTTTACTGGCACAGCCAAAATCT 1209

 265 LeuLeuSerGlyGlyLeu...................ValProTh 273
     |||                                  ::
1210 TTGATTGATGANTAACACCAATACCGCACCTTTTGTTAAAGAGGTTGTCTC 1259

 273 rLysProGlnThrProGlyAspProGlnProCysAsnGlnProGlnInT 290
     ::||||||:::||||
1260 CAAAAAAGCCAATCCCAACAACCCA.......GACCCCAAC 1293

 290 hrThrSerValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysLeu 306
     :::                               |||
1294 ...............................TCAGATACG 1302

 307 LeuGluGlyAlaThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGl 323
     |||||||||:::                      :::||
1303 CTAGAAGGCGGGGGTTTTATGGTGGAGTCGGGCCGANGAGCTG. 1341

 323 nAlaArgValPheSerSerAsnAspIle..................G 333
     |||   ||||||||                       |
1342 .GCGGGTAAAATTTTATCCAATGACAACGGCAACTTTGTGGGTCTTTGGTG 1390

 333 lyGluArgIleGluLeuSerAspGlyThrTyrThrLeuThrGluAsn 349
     ::||||||        |||
1391 GCAAACGAGACAAAACGACCGAACCTGTGCCCACAAAAACGGTGTAGTTT 1440

 350 SerProAlaGlyThrTyrSerIleAlaGluProIleThrPheLysValGlu 366
     :::  ||||                    |||
1441 AGT.....ACAGGATTGAA. 1481

 366 aGlyLysValTyrThrIleIleAspProLysLys..........GlnIleG 379
     :::::|||||||                         |||
1482 TGAAGAGATTGGTAGCATTATTGACGGTAAAAGTAAATGATGAAGTCA 1531

 379 luAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAsn 395
     |||                              |||
1532 ATAATCAAATTGAAGATGAAACTGTCCCTGTCAGTAATAAGAATATTAT 1581

 396 AspPhe............GluGluPheSerVa 402
     :::::::              ::::::::
1582 GAATATAATTATTGGACGACCCAACAAACAATTCACCAAAAAAATAAAGGC 1631
```

```
 402 lLeuThrThrGlnAsnTyrAla...............LysPheTyrT 413
     ::  :::|||                           ||||||||||
1632 CAGGGTCCAAAAAAACCCTGCTTATTTTGGTCAGCATGATAAGTTTTATT 1681

 413 yr............AlaLys.....AsnLysAsnGly 419
     :::              |||||       |||
1682 TTAATNGGTAACTATTATTGACTTATCAGCCAAAGAAGCAAACAACAGAC 1731

 420 SerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProProAs 436
     |||||||                             :::|||
1732 GTCTCCCAAGATACCAGCACCAATAAAGAGTATTTTGGCTAAAAATACCCAGA 1781

 436 pSerGluAspGlyGlyLysThrMetThrProAspPheThrThr 450
     |:::                    |||      :::|||
1782 TGCC.......AAAGTAAGCACAGCAATAAAAAGTTACCAAAAATCG 1819

 450 ...........GlyGluValLysTyrThrHisIleAlaGlyAr 461
                 ||              :::||
1820 TTCTACAACAAGCCAAAAGATAAGCCGTATACCGGCCATTCANTGCCAAAAGC 1869

 461 gAspLeuPheLysTyrThrValLysProArgAspThrAspProAspThrP 478
     ::|||                             :::|||
1870 TATGACCACCATCACTTTTGGTGAAGTATTGATTGATAATAACAAAGGCAA 1919

 478 heLeuLys.................................His 481
     |||                                     :::
1920 CCCAACACGCCAGTTATTTTGTGCAAAGGCGGGTCAAGCGGATGTCAGTACTC 1969

 481 ...........GlyGluValLysTyrThrHisIleAlaGlyLe 498
                 ||              :::||
1970 AGCTGCCCAGTGCAGGTAAAATTCACCTATAATGGTCTTTGGGCAGGCTAC 2019

 498 uTyrSerGlyLeuThrGluThrGlnInLeuArgAlaAlaThrGlnLeuAlaI 515
     :::|||||                          |||
2020 CTGACCCAGAAAAAAGCACAAAGGTTATNCCANAGAATGAGGATACCATCAA 2069

 515 leTyrTyrPheThr........................ 519
     :::
2070 GCAAAAAGGTCTTAAAGAT. 2088

 519 ...........TyrHisGlyPh 535
                 :::
2089 ..TATATATTGACCAAAGACTTTATCCCACAAGATGACGATGAC 2136

 535 eGlyAspMetAsnAspSerThrLeuAlaValAlaLysIleLeuValGlu 552
     |||||||                     |||
2137 GATAGTTTGACCCGCCATTCTGATGATTCACAAGANTGATAATAACACATGGCCA 2186

 552 yrAlaLeuGlnAspSerAsnProGlnLeuThrAspLeuAspPhePheIle 568
     :::                   ||| |||  :::
2187 TGATGCATTTG.......ATTGCATCTGANG 2209

 569 ProAsnAsnAsnLysTyrGlnSerLeuIleLeuGlyThrGlnTrpHisProGl 585
     :::                              |||
2210 ATTCACAGATGATGACACAGATGGCGGATGACGAATTCAGATGATTTG... 2256

 585 uAspLeuValAspIleIleArgMetGluAspLysLysGluValIleProV 602
     |||||                         :::|||
2257 ..GGTGATGGTGCAGATGATGAGCGCCGCAAAGTGTATCATGCAGG 2303

 602 al..............ThrHisAsnLeuThrLeu..... 608
     :::              |||||::::
2304 TAAT............ATTCGCCCTGAATTNGAAAAACAAATACTTGCCCA 2341

 609 .....................ArgLysThrValThrGlyLeuAlaGl 617
                           :::::||||
2342 TTAATGAGCCTACTCATGAAAAAAACCTTTGCCCTAGANGGTAAAAATAAG 2391

 617 yAspArgThrLysAspPheHisPheGlu...GluLeuLysAsnLysG 634
     ::::|||                        :::
2392 GCTAAGTTTTGATGTAAACTTTGACACCAACAGCCTAACTGGTAAATTAA 2441
```

```
                                        Quality:    161.00        Length:  755
                                        Ratio:       0.434        Gaps:     40
                                        Percent Similarity: 49.139    Percent Identity: 21.060

alignment_block:
US-09-494-297-2 x X20201

Align seg 1/1   to: X20201     from: 1   to: 4249
US-09-494-297-2 from: 1   to: 4249
```

seq_documentation_block:
```
ID   X20201 standard; DNA; 4249 BP.
XX
AC   X20201;
XX
DT   20-APR-1999 (first entry)
XX
DE   Enterococcus faecalis EF104 gene fragment.
XX
KW   Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW   detection; attenuation; antigenic; ss.
XX
OS   Enterococcus faecalis.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850554-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998.
XX
PR   14-NOV-1997;  97US-0056009.
PR   06-MAY-1997;  97US-0044031.
PR   16-MAY-1997;  97US-0046655.
XX
PT   New isolated Enterococcus faecalis polynucleotides - used to develop
PT   products for the detection of Enterococcus and for use in vaccines
PT   for prevention or attenuation of Enterococcus infection
XX
PS   Claim 1; Page 202-203; 301pp; English.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
DR   WPI; 1999-070095/06.
DR   P-PSDB; Y00211.
XX
CC   The present sequence encodes an antigenic polypeptide fragment
CC   isolated from Enterococcus faecalis. The present invention describes
CC   genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC   The proteins can be used in vaccines for preventing or attenuating an
CC   infection caused by a member of the Enterococcus genus in an animal.
CC   They can also be used for detecting Enterococcus antibodies in a sample.
CC   The nucleotide sequences can be used for detecting Enterococcus nucleic
CC   acids. Products from the present invention can also be used for
CC   screening compounds to identify agonists and antagonists of E. faecalis
CC   protein activity.
XX
SQ   Sequence 4249 BP; 1579 A; 779 C; 815 G; 1069 T; 7 other;
```

seq_name: /cgn1_8/gcgdata/geneseq/NA1999.DAT:X20201

alignment_scores:

```
2442 CGATGAGAGGAGGTTGAATATCGTCTTTGATATC.....AAAAATGGCAAAA 2485
     |||:|:|:                  ||||||:|||
2486 TTGATGGCACAGGATTTACCGCCAAAGCCGATTGCCAAACTATCGTGAA 2535

634 lnGluLeuLeuSerGlnThrValLysThrAspLysThrAsnLeuGluPhe 650
    |||:|:
2486 ...

651 LysAspGlyLysAlaThrIleAsnLeuLysHisGlyGluSerLeuThrIle 667
    |||
2536 GAAGTGGGT..........AACAACCAAGGTGGC........ 2559

667 uGlnGlyLeuProGluGlyThrTyrSerTyrLeuValLysGlyGluThrAspSerG 684
    |||:||           |||||:::
2560 ...............GGTTTCTTATACAACACATCAAAGATAATTCGATGTTA 2593

684 luGlyTyrLysValLysValAsnSerGlnLysGluValAla 696
    :|||
2594 AGGGGCAANTTTTTTGGCACAAANTGGCGAAGAGTTGGCA 2631
```

```
1548 AAACTGCTTGGGGTAATTATGATCAA.                        1573
     ||:::::||
  39 AlaLeuValThrSerMetValGlyAlaLysThrValPheGlyLeuValGl  55
1451 TCATTAAGTACACCTGTAATTGGTTCTAATAAAGCCATTCAATTACTATC 1500

  55 uSer........  SerThrProAsnAlaIleAsnProAspSerSerSerG  69
       |||            ::: ||| :::    |||      ||||
1501 CGATCAATATATTGAACCAATTAGTGTGTTGTTGTAAATCCT...TTGAATGCTG 1547

  69 luTyrArgThrTrpTyrGlyTyrGluSerTyrTyrValArgGlyHisProTyrTyr  85
       |:::|||  |||:::||      |||
1574 ........AATGGTGCCTATTCATCAAGAACAACTGTCTCAGTTANGGG 1614

  85 .........                                          
1574 ........                                           

  86 LysGlnPheArgValAlaHisAspLeuArg........ValAsnLeuGluGl 100
         |:|                 |||               |||    |||
1574 ........AATGGTGCCTATTCATCAAGAACAACTGTCTCAGTTANGGG 1614

 100 ySerArgSerTyrGlnValTyrCysPheAsnLeuLeuLysLysAlaPhePro.  116
      |||||:::  ||            |::::|||          |||
1615 AAGCAAAGAGAAAACCGATTCAAAATTAGAAATTAAAGTAAAGCAATCCTA 1664

 117 .........  LeuGlyLysAspSerSerValLysLysLysTrpTyrLysLys 129
            |||     ::: |||::|||     ||| ||::||||
1665 ATTATCTTTCATTACGAGCTACAAAAGAAAATTTATTTTTATTACAG... 1711

 130 HisAspGlyIleLeuSerThrLysPheGluAspTyrArgAlaMetSerProArgIle 146
        :::||:     ||:::||      |:::||:|  :||
1712 .........TTAGGAACG.....GATTATACAGTAACGCCAACGTC 1743

 146 eThrGlyIleAspGluLeuAsnGlnLysLeuArgAlaValMetTyrAsnGlyH 163
       ||||             |||     :::|||  |||
1744 AGATGGTTCAGTTATTAAGTTCACTACGCCAATAACCAACGAAATCCAAA 1793

 163 isProGlnAsnAlaAsnGlyIleIleMetGluGlyLeuGlu.........  175
        |||     |||::: |||   |||:::||||
1794 TTCCAATTGGTTTTAATTATGTCGCAGATAGTTTGCCAAAAGAATAAAAGT 1843

 176 ...ProLeuAsnAlaIleAlaArgValThrGlnGluAlaValTrpTyrTyrSe 191
          ||:::::||||       |||    :::||:|||  :::
1844 ATCCCAGTCGATACGATACCGATAACAAATGAGTCGTGAAGGTTTAACTCC 1893

 191 rAspAsnAlaProIleSerAsnProAspGluSerPheLysArgGluSerG 208
       ||::: ||          |||  |||     |||||   |||
1894 AGTTGATACGACAGTAACTACT..........AATAGTAAGCGTGGTTCTG 1934

 208 luSerAsnLeuValSerThrSerGlnLeuSerLeuMetArgGlnAlaLeu 224
       ||        ||          |||   |||||||||
1935 AACGAACACACTTCAAAGTAGTAAAAATCAAATTCCTTGTCAATGCACGAAAT 1984

 225 LysGlnLeuIleAspProAsnLeuAlaThrLysSerMetProLysGlnValPr 241
       |||||||        ||| :::||||
1985 GATTCTTTTGACTCACTCAAGCGTCCGTACAAAAATTCCAGCTGGCGCC. 2032

 241 oAspAspPheGlnLeuSerIlePheGluSerGluAspLysGlyAspLysT 258
      ...GAATGTCTTTTGACATTTTATGATGTTTCAAAC..... .GATCAGG 2072

 258 yrAsnLysGlyLeuTyrGlnAsnLeuLeuSerGlyLyrGlyLeuValProThrLys 274
       |:::         ||| :::|||    |::: ||||
2073 TAGATTCAATTTATCCACAATACTGGGACCCGCGGTCAAATACTTTGATAAA 2122

 275 ProProThrProGlyAspProMetPro.........ProAsnGlnProGl 289
       ||| |||||||||||                    ||| :::
2123 CCAATGACGGCAAACAGCCCCTGGATANCCACGATTACTTTTGACGAAA 2172
```

```
 289 nThrThrSerValLeuIle.............................ArgLysTyrAlaI  300
     |||:||::|||                                    :||              |
2173 TACCAATAGTTACACGGTTGATTTGATTTGGAAAAACCAACAAACGTTACATTA        2222

 300 leGlyAspTyrSerLysLeu........................LeuGluGlyValAThrLeuGln  313
     ||:                                         ||||||
2223 TT...GAGTATAAAAACCCAATGGCTCGGATCGACGTGCCAACTCTTTAT              2269

 314 LeuThrGly...........................AspAsnValAsnSerPheGlnAl  324
     ||                                 :::||||||
2270 ATAACAGGGACAACGGAAAGACACCAACAATCGAATAATAATGAAGGCTCTGC          2319

 324 aArgValPheSerAsnAspIleGlyIleGluLeuSerAspG  341
     |:|                                     |||
2320 TTCCGGTTTCTGTTCAAAAT.......GAACGGTTAGACATTTGAGTG              2360

 341 lyThrThrTyrThrLeuThrGluLeuAsnSerProAlaGlyThrTyrSerIleAla  357
     :||                                                   |||
2361 CAACA.......................................CAAGCGCGCG      2374

 358 GluProIleThrPheLysValGluValGlyLysValTyrThr.....IleIl  373
     ||:||                                         ||     ||
2375 AATCCACACATTAAAAAAGTAACACACACAGTAACACAAAAAAT                   2424

 373 eAspGlyLys.........GlnIleGluAsnProAsnLysGluIleValGluP  388
     :||:|||                                             ||
2425 TGATAATAAAAACACATCGTGTGAAAAATCCACACGATTTGAATTAACACCAA         2474

 388 roTyrSerValGluValGlyTyrAsnAspPheGluSerValLeuThr  404
     ||||                                          |||
2475 AAGGCACACAACCAATGCTCAAAATCGATTTGAATTCTATTACCGTG                2518

 405 ThrGlnAsnTyrAlaLysPheTyrTyrArgTyrAlaLysAsnGlySerSe  421
     |||||||                                          |||
2519 ............................................AAAGGCGTGCC        2529

 421 rGlnValValTyrCysPheAsnAlaAspLeuLysSerProProAspSerG  438
     :||                                             |||
2530 AGAAGATGCTTAT...................................TCATTAGAGAAGACTA 2558

 438 luAspGlyLys.........GlnIleGluAsnProAspPheThrThrGlyGlu  452
     :|||:|||                                            |||
2559 CAAACGGTGCGAAAGTCATTTTTAAAGACTATACATTGACAGAAAAACATT          2608

 453 ..ValLysTyrThrHisIleAlaGlyArgAspLeuPheLysTyrThrVa  468
     :::|||::|||                                     |||
2609 ACGATTGAATACAAATACCGTCTCTGCAAACGCTGGCCAAAATCTATACAGA         2658

 468 LLysProArgAspThrPheLeuLysHisIleLeuLysSV  485
     ||||||                                 ::|||
2659 AACAACAAATCGACTCTGAAACAATTGAACCAGATGTCTCGCTAGCAAGAAAA        2708

 485 aIle.................................GluLys  488
     :::                                      ::
2709 AAGTCACCACTGCGCCAATCACATTGAAATTCTTCAGAAGGTGATGCGGAA          2758

 489 GlyTyrArgGluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGluTh  505
     |:|                                             ||||||
2759 GGTATTGTTTATTTAGCAACTGCCACATTCTCACGCATAACGTAGAGA             2808

 505 rGlnLeuArgAlaAlaThrGlnLeuAlaIleGluTyrTyrPheThrAsp....  520
     ::|:                                           ::
2809 TGAAAAACCAAGCAAATTGCGAAGGTTTCTTTTGAACTAATTGATAATGTCA         2858

 521 .........SerAlaGlyLeuAspLysAspLeuLysAspTyrHisGly  534
     :::|||||                                       |||
2859 CGCCATACAGCACCGGAATTTACAACAGATGAAAAAGTCAAAtaCTCCTTT          2908

 535 PheGlyAspMet...AsnAspSerThrLeuAlaValAlaLysIleLeuVa  550
     |||            :::|||                            ||
2909 GATGCCATCATCGACAGGTGATTATACTTTGCGAGTAACGAAATGTACCGCA         2958

 550 lGluTyrAlaGlnAspSerAsnProProGlnLeuThr.............  562
                                                           
```

(column 2)
```
2959 GGAATATTCCGTTGGATGAGACAG..............TATTTGACAGGAAAAAGCCATTA 3002
     ||||||:::                              ::||||
 563 .............AspLeuAspPhePheIlePro.............. 569
     ||||||                                ||||||
3003 AGCTGGTCAAAGGAGGAGACCAACCAACTAAAAAAATTCCATTAACGAAAAACAATT 3052

 570 AsnAsnAsnLysTyrGln...............SerLeuIleGlyThrGlI 581
     :::|||                             :::||
3053 GATCACAGTCGTTTACAAGTCAAAAGATTCAAACGATTTATCTCGGCGATTC 3102

 581 nTrpHisProGluAspLeuValAspIleIleAlaArgMetGluAspLysLys. 597
     |||                                            ||||||
3103 ATGGAAACCAGAAGAG......AACTTTGTTTCAACAACAGATAAAACAG 3146

 598 .GluValIleProValThrHisAsnLeuThrLeuArgLysThrValThr 613
     :||                                            ||||||
3147 GTCAAGACGTTCCCTTCGAA..............AAAATCACGTTTCA 3181

 614 GlyLeuAlaGlyAspArgThrLysAspPheHisSerPheGluIleGluLeuu 630
     |||                                              ::|
3182 GGTCAAGTTGATACACNCAAAAGCAGGCGTTTATCCAAATTATTTACAGTGA 3231

 630 sAsnAsnLysGlnGluLeuLeuSerGlnThrValLysThrAspLysThrA 647
     :::|||                                          |||
3232 CGAAGGTAAAGAAGAAAACACCCTAGTGACCGTCAAACCCGACCAATCTA 3281

 647 snLeuGluPheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGlu 663
     ::|||||                                          ||:::
3282 AGTTAGAGGTGCAAAGAT......ACAACGATTTATAGTT...........GGTGAT 3319

 664 SerLeuThrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysSI 680
     |||                                             :::|||::
3320 TCGTGGAAA....................CCAGA......GATAAATTTCGTTTCAGC 3351

 680 uThrAspSerGlyGluGlyTyrLysVal..........LysValAsnSerGlnG 694
     ||||||                                ||| |||
3352 GACAGCACAAAAACAGGTGCAAGACGGTCAGACGGTNCCGTTTGAAAAAATTGATGTTGCAGG 3401

 694 luValAlaAsnAlaThrValSerLysThrGlyIleThrThrSerAspGluThr 710
     ::|||||||                                        |||
3402 GAACAGTGAAT.......GTTGATAAAATAGGC..........GATTATGAA 3436

 711 LeuAlaPheGluAsn 715
     ::::::::::|||
3437 ATTGTCTATAAAAAT 3451
```

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
DR WPI; 1999-070095/06.
DR P-PSDB; Y00210.
XX
PT New Isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection .
XX
PS Claim 1; Page 200-201; 301pp; English.
XX
CC The present sequence represents a gene isolated from
CC Enterococcus faecalis. The present invention describes genes, proteins
CC and antigenic polypeptides isolated from E. faecalis. The present sequence can
CC be used in vaccines for preventing or attenuating an infection caused
CC by a member of the Enterococcus genus in an animal. They can also be
CC used for detecting Enterococcus antibodies in a sample. The nucleotide
CC sequences can be used for detecting Enterococcus nucleic acids. The
CC products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of E. faecalis protein
CC activity.
XX
SQ Sequence 4359 BP; 1614 A; 794 C; 838 G; 1106 T; 7 other;
XX

alignment_scores:
   Quality: 161.00
   Ratio: 0.434
   Percent Similarity: 49.139   Percent Identity: 21.060

alignment_block:
us-09-494-297-2 x X20200

Align seq 1/1   to: X20200   from: 1   to: 4359

                                        Length: 755
                                        Gaps: 40

```
39   AlaLeuValThrSerMetValGlyAlaLysThrValPheGlyLeuValGl   55
     ::|||||    ||      ||    ::  ||||||:   :||
1546 TCATTAGTACACCTGTAAATTGGTCCTAATAAAAGCCAATCAATTAGTATC  1595

55   uSer........SerThrProAsnAlaIleAsnProAspSerSerSerG   69
       ::::        ::::::||||||||:|||||||||
1596 CGATCAATATATTGAACCAATTAGTGTTGTTGTTAATCCT...TTGAATGCTG  1642

69   luTyrArgTrpTyrGlyTyrGlyLeuSerTyrValArgGlyHisProTyrTyr  85
        ||  |   ::|||||::|::||::
1643 AAACTGCTTGGGGTAATTATGATCAA  1668

86   LysGlnPheArgValAlaAlaHisAspLeuArg.....ValAsnLeuGluGl  100
     ||||:||   |||||    :::||||   |  ||
1669 .........AATGGTGCCTATTCATCAAGAACAACTGTCTCAGTTATGG  1709

100  ySerArgSerTyrGlnValTyrCysPheAsnLeuLysLysAlaPhePro.  116
       ||||:   :::::::||||||:|   ||
1710 AAGCAAAGAGAAACCGATTCAAAATTTAGAAATTAAAGTAAAGCATCCTA  1759

117  .........LeuGlySerAspSerSerValLysLysTrpTyrLysLys  129
     ||        ||   :::|||||::|||||
1760 ATTATCTTTCAATTACGAGCTACAAAAGAAATTATTTTATTACAAG...  1806

130  HisAspGlyIleSerThrLysPheGluLysAspThrTyrAlaMetSerProArgIl  146
     :::||||||    ||      ||:   ||||:||||   |||
1807 .........TTAGGAACG......GATTATACAGTAACGCCAACGTC  1838

146  erThrGlyAspGluLeuAsnGlnLysLeuArgAlaValaMetTyrAsnGlyH  163
     |||:|   |   ||:::|||||
1839 AGATGGTTCAGTTATTAAGTTCACTACGGCCCAATAACCAACGAAATCCAA  1888

163  isProGlnAsnAlaAsnGlyIleMetGluGlyLeuGlu..........  175
     |||  :|    ||    ||   :::||||
1889 TTCCAATTGGTTTTAATTATGTGCCAGATAGTTGTTGCCAAAAGATAAAAGT  1938

176  ...ProLeuAsnAlaIleLeuArgValThrGlnGluAlaValaValTrpTyrTyrSe  191
        :||:||||    ||:||     ||||||:|||:|:||
1939 ATCCCAGTTGGATTACGGATACCGATAACCAATACCAATGAGTGCTGAAGGTTTAACTCC  1988

191  rAspAsnAlaProIleSerArgProAspGluSerPheLysArgGluSerG  208
       ||||||     |||     ::::    |||||||
1989 AGTTGGATAACGGACAGTAACTACT.......AATAGTAAGCGTGGTTCTG  2029

208  luSerAsnLeuValSerThrSerGlnLeuSerLeuMetArgGlnAlaLeu  224
     ||    :::||    |||   :   |||||   ::||
2030 AACGAACACTTCAAAGTAGTAAAAAATCAATTCCTTTGTCAATGCACGAAAT  2079

225  LysGlnLeuIleLeuAspProAsnLeuAlaThrLysMetProLysGlnValArg  241
     |||   |||||    :::    ||||||||:|||||||
2080 GATTCTTTTGACTCACTCAAGCGTCCGTACAAAAAAATTCAGCTGGCGCC.  2127

241  oAspAlaPheGlnLeuSerIlePheGluSerGluAspSerGlyAspLysGlyLysT  258
      ::::::::||    ||||||||||||||:|   :::
2128 .GATGTTCTTTTTGACATTTTATGATGTTCAAAC.......GATCAGG  2167

258  yrAsnLysGlyLeuTyrGlnAsnLeuLeuSerGlyGlyLeuValProThrLys  274
       :::||    ||      :::||||   |||
2168 TAGATTCAATTTATCCAACAACTATCTGGGAACCGGGTCAATACTTTGATTAAAA  2217

275  ProProThrProGlyAspProProMetPro.......ProAsnGlnProGl  289
     |||        ||||   :::        |||
2218 CCAATGACGGCCAAACAGCCCTGGAAATCCAACGATTACTTTTGACGAAAA  2267

289  nThrThrSerValLeuIle..........ArgLysTyrArgAlaI  300
     :::||:||||||          |||:::|||
2268 TACCAATAGTTACACGTTTGATTTTGGAAAAACCAACAAAAGGTCTCTGC  2317

300  leGlyAspTyrSerLeuSer.........LeuGluGlyAlaThrLeuGln  313
     |||        :::         |||||||||:|:
2318 TT...GAGTATAAAAACGCCAAATGGCTGGATCGACGGTGCCAAACTCTTTAT  2364

314  LeuThrGly...........AspAsnValAsnSerPheGlnAl  324
     ||||||           :::|||||||||
2365 ATAACAGGGACAGCAGCAAAAGAAAACCACAATGCAATCGAATTAATGAAAAT  2414

324  aArgValPheSerSerAsnAspProIleGlyGlyValArgIleGluGluLeuSerAspG  341
     ||||||        :::||||||||||||
2415 TTCGGTTTCTGTTCAAAAT.......GAAGCGTTAGACATTTTGAGTG  2455

341  lyThrHisTyrThrLeuThrGlyLeuAsnSerProAlaGlyTyrTyrSerIleAla  357
     ||        :::||||   |||||
2456 CAACA.............CAACGCGCG  2469

358  GluProIleThrProPheLysValGluAlaGlyLysValTyrThr...IleIl  373
     |||        ||||:||    :::    |||
2470 AATCCAACATTAAAAAAATGTAACAAAAACGACAGTAACAACAAAAAATAT  2519

373  eAspGlyLys.....GlnIleGluAsnProAsnLysGlyGluIleValGluP  388
      ::|         |||||||||:|||||||
2520 TGATAATAAAACACATCGTGTGAAAAAATCCAACGATTGAATTAACACCAA  2569

388  roTyrSerValGlyAlaTyrAsnSerPheGluSerProSerValLeuThr  404
      :::||||||     ||||||   |||
2570 AAGGCACAACCAAATGCTCAAATCGATTTGAATTCTTATTACCGTG...  2613

405  ThrGlnAsnTyrAlaLysPheThrTyrAlaLysSerAsnGlySerSe  421
     ||||        :::||||||   |||
2614 .........GCAATCAATTATATAAGCAAAGTCTTTTAACGTTAT.  2624

421  rGlnValValTyrCysPheAsnAlaAspSerLeuLysSerProProAspSerG  438
       :::|         ||:||
2625 AGAAGAATGCTTAT..........TCATTAGAAGAGACTA  2653

438  luAspGlyIleGlyLysThrMetProAspPheThrThrGlyGlu......  452
     :::|||||:||||||:|||||
2654 CAAACGGTCGCAAAGTCATTTTTTAAAGACTATTACATTGACAGAAAACATT  2703

453  ...ValLysTyrThrHisIleAlaGlyArgAspLeuPheLysTyrThrVa  468
     :||||:|:|:|||
```

```
2704 ACGATTGAATACAAATACGGTCTCTGCAAACGCTGGGCCAAATCTATACAGA 2753
       :||||:::
        :::
 468 LLysProArgAspThrAspProAsnProThrPheLeuLysHisIleLysLysV 485
     ||:|:::

2754 AACAACAATCGACTCTGAAACATTGAACCAGATGTCTCCTAGCAAGAAAA 2803
      |||||
 485 alIle............Glulys 488
       :::

2804 AAGTCACCACTCGCGCCAATCACATTGAAAATTCTCAGAAGGTGATGCGGAA 2853
       :::
2904 TGAAAACCAAGCAAATTGCGAAGGTTTCTTTTTGAACTAATTGATAATGTCA 2953
       :::

 489 GlyTyrArgGluLysSGlyGlnAlaIleGluTyrSerGlyLeuThrGluThr 505
       |||    :|||
 505 rGlnLeuArgAlaAlaThrArgGlnLeuAlaIleLeuTyrTyrPheThrAsp 520
      :::|||

2854 GGTATTGTTATTTAGCAACTGGCCACATTCTACACGCATAACGTAGAGGA 2903
       |||
2954 CCCATACAGCACCACCGAATTTACAACAGATGAAAAAGGTCAATACTCCTTT 3003
       |||:::

 521 ......SerAlaGluLeuAspLysAspLysLeuLysAspTyrHisGly 534
       :||:::||
3004 GATGCCAATCATGACAGGTGATTATACTTTGCGAGTAACGAAGTACCCGCA 3053
       |||

 535 pheGlyAspMet...AsnAspSerThrLeuAlaValAlaIleLeuLeuVa 550
     :::|||
3054 GGAATATTCCGTGGATGAAGAG......TATTTGACAGGAAAAGCCATTA 3097
      :::|||:::

 550 lGluTyrAlaGlnAspSerAsnProProGlnLeuThr 562
     |||||:::
3098 AGCTGGTCAAAAGAGACGACAACCAACTAAAAATTCCATTAACGAAAACAATT 3147
       |||

 563 ......AspLeuAspPhePheIlePro 569
       |||:::
3148 GATCACACGTCGTTACAAGATCAAAGATTCAACGATTTATGTCGGCGATTC 3197
       :::|||

 570 AsnAsnAsnLysThrYrGln 581
       :::|||
 581 nTrpHisProGluAspLeuValAspIleIleLeuArgMetGluAspLysLys 597
     :::||||
3198 ATGGAAACCAGAGAG......AACTTTGTTTCAGCAACAGATAAAACAG 3241
      :::|||||

 598 .GluValIleProValThrHisLeuAsnLeuArgLysThrValThr 613
       :::   :||||||
3242 GTCAAGAGGTTCCTTCGAA......AAAATCACTGTTTCA 3276
       :::|||

 614 GlyLeuAlaGlyLysAspArgThrLysAspAsppheHisPheGluIleGluLeuy 630
       |||   :||||
3277 GGTCAAGTTGATAACANCAAGCAGGCGGTTATCCAATTATTTACAGTGA 3326
     :::||||::

 630 AsnAsnLysGlnGluLeuLeuSerGlnThrArgValLysThrAspLysThrA 647
     :::|||||||||
3327 CGAAGGTAAAGAAGAAAAACGCCTATGTGACCGTCAAACCCGACCAAATCTA 3376

 647 snLeuGluPheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGlu 663
     :||||:::||||
3377 AGTTAGAGGTCAAAGAT......ACAACGATTTATGTT.....GGTGAT 3414
       :::||||:::

 664 SerLeuThrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysGl 680
       ||||
3415 TCGTGGAAA......CCAGAA......GATAATTCGTTCAGC 3446

 680 urThrAspSerGluGlyTyrTyrLysVal......LysValAsnSerGlnG 694
     ||||||:::
3447 GACAGACAAAACAGGTCAAGAGACGGTNCCGTTGAAAAAATTGATGTTCAG 3496

 694 luValAlaAsnAlaThrValSerLysThrGlyIleThrSerAspGluThr 710
       :||:::|||
3497 GAACACGTGAAT......GTTGATAAAATAAGGC 3531

 711 LeuAlaPheGluAsn 715
       :::|||
```

alignment_scores:
```
          Quality:  161.00              Length:    755
          Ratio:    0.434               Gaps:       40
Percent Similarity:  49.139    Percent Identity:  21.060
```

alignment_block:
US-09-494-297-2 x X13179

Align seq 1/1 to: X13179
US-09-494-297-2   from: 1   to: 12445

```
  39 AlaLeuValThrSerMetValGlyAlaLysThrValPheGlyLeuValGl 55
     ::|||  |  |::::::|:||||  :  :::|||::
9634 TCATTAAGTACACCTGTAATTGGTCCTAATAAAAGCCATTCAATTAGTATC 9683
```

```
55  user........SerThrProAsnAlaIleAsnProAspSerSerSerG  69        341  lyThrTyrThrLeuThrGluLeuAsnSerProAlaGlyTyrSerIleAla  357
            :::||||                    :::|||                           10503  TTCGGTTCTCGTTCAAAAT........GAAGCGTTAGACATTTGAGTG  10543
9684  CGATCAATATATATTGAACCAATTAGTGTTGTTTAATCCT...TTGAATGCTG  9730

69  luTyrArgTrpTyrGlyTyrGluSerTyrValArgGlyHisProTyrTyr  85        357  ...CAACA.  357
      ||:|||||                                                       10544  CAACA........CAAGCGGCG  10557
9731  AAACTCCTTGGGTAATTATGATCAA  9756

86  LysGlnPheArgValAlaHisAspLeuArg......ValAsnLeuGluGl  100       358  GluProIleThrPheLysValGluAlaGlyLysValTyrThr...IleIl  373
            |||:::::            |:|                                    |||:||||            :|                    ||
9757  .......AATGGTGGCTATTCATCAAGAACAAACTGTCTCAGTTATGG  9797       10558  AATCCAACACATTAAAAATGTAACAAAAACGACAGTAACAACACAAAAATAT  10607

100  ySerArgSerTyrGlnValTyrCysPheAsnLeuLysLysAlaPhePro.  116      373  eAspGlyLys........GlnIleGluAspAsnLysGlyIleLeuValGluP  388
       ||||:::::                          :::|||                    |||                 :::::::|||||||::::
9798  AACCAAAGAAGAAAACCGATTCAAAATTTAGAAAATTAAAGTAAGCATCCTA  9847      10608  TGATAATAAAACACATCGTGTGAAAAATCCAACGATTGAATTAACCACCAA  10657

117  ........LeuGlySerAspSerSerValLysTrpTyrLysLys  129          388  roTyrSerValGluValAlaTyrAsnAspPheGluGlupheSerValLeuthr  404
                 |||||:::::||||||                                    :::|||            |||          :|
9848  ATTATCTTTCATTACGAGCTACAAAAAGAAATTTATTTTATTACAG  9894          10658  AAGGCACAACCAAATGCTCAAATCAGATTTGAATTTCATTACCGTG........  10701

130  HisAspGlyIleSerThrLysPheGluAspTyrAlaMetSerProArgIl  146      405  ThrGlnAsnTyrAlaLysPheTyrTyrAlaLysAsnLysAsnGlySerSe  421
                       :::|||                                                              :::|||
9895  .......TTAGGAACG  9926                                          10702  ........TCATTAGAGAGACTA  10712

146  eThrGlyGluLeuAsnGlnLysLeuArgAlaValMetTyrAsnGlyH  163        421  rGlnValValThrCysPheAsnAlaAspLeuLeuLysSerProProAspSerG  438
                 |||                  :::||                                |||                 :::|||
9927  AGATGGGTTCAGTTATTAAGTTCACTACGGCCAATAAACCACGAAAATCCAA  9976     10713  AGAAGATGCTTAT........AAAGGCGTGCC  10741

163  lsProGlnAsnAlaAsnGlyIleMetGlyGlyLeuGlu.........  175        438  luAspGlyGlyLysThrAlaMetThrProAspPheThrThrGlyGlu...  452
            |||                         |||                              |||                  :::|||
9977  TTCCAATTGGTTTAATTATATGTTGCCAGATAGTTGCCAAAAGATAAAAGT  10026     10742  CAAACGGTGCGGAAAGTCATTTTTAAAGAACTATACATTGACAGAAAACATT  10791

176  ...ProLeuAsnAlaIleAlaArgValThrGlnGlnValAlaValTrpTyrTySe  191  453  ...ValLysTyrThrHisIleAlaGlyArgAspLeuPheLysTyrThrVa  468
      |||:||||||::|||                  ||||                          |||:||||||                  ||||||
10027  AtCCCAGTCGATACGAATACGATAACGATGTGCTAGAAGGTTTAACTCC  10076       10792  ACGATTGAATACAATTACGGTTCTCGCAAAGCTGGCCAAATCTATACAGA  10841

191  rAspAsnAlaProIleSerAsnProAspGluSerPheLysArgGluSerg  208      468  lLysProArgGlyAspThrPheLeuLysHisHisIleLeuLysLysV  485
            |||                  :::                                      |||                  ||||||
10077  AGTTGATACGACAGTAACTACT........AATAGTAAGCGTGGTTCTG  10117       10842  AACAACAATCGACTCTGAAAACATTGAACCAGATGTCTGCTAGCAAGAAA  10891

208  luSerAsnLeuValSerThrSerGlnLeuMetArgGlnAlaLeu  224          485  alIle.......GluLys  488
      ||                  :::|||                                             :::
10118  AACGAACACTTCAAAGTAGTAAAAATCAAATTCCTTGTCAATGCACGAAAT  10167     10892  AAGTCACCACTGCGCCAATCACACATTGAAATTCTCGAGAAGGTGGATCGGAA  10941

225  LysGlnLeuIleAspProAsnLeuAlaThrLysMetProLysGlnValPr  241      489  GlyTyrArgGluLysGlyGlnAlaIleLeuTyrSerGlyLeuThrGluth  505
      ||||||||::|||                                                    |||                  :::|||
10168  GATTCTTTTGACTCACTAAGCGTCCGTACAAAAAATTCCAGCTGGCGCC..  10215     10942  GGTATTGTTTATTAGCAACTGCCACATTCACGCCATAACGTACGGCA  10991

241  oAspAspPheGlnLeuSerIlePheGluSerGluAspLysGlyAspLyst  258      505  rGlnLeuArgGlnAlaAlaThrGlnLeuAlaIleLeuThrTyrTyrPheThrAsp  520
            :::                                                          rGlnLeu                  :::
10216  .....GANGTTCTTTTTGACATTTATGAGTTTCAAAC.......GATCAGG  10255      10992  TGAAAAACCAAGCAAATTGCGAAGGTTTCTTTTGAACTAAATTGATAATGTCA  11041

258  yrAsnLysGlyGlyTyrGlnAsnLeuLeuSerGlyLeuValProThrLys  274      521  .......SerAlaGluLeuAspLysAspLysLeuLysLysAspTyrHisGly  534
            :::|||                                                                |||||:::::
10256  TAGATTCAAATTTATCCACAATACTGGGACCGGGTCAATACTTTTGATAA  10305      11042  CGCCATACAGCAACCGAATTTTACAACAGATAGAAAAGGTCAATAACTCCTTT  11091

275  ProProThrProGlyAspProProMetPro.......ProAsnGlnProGl  289      535  PheGlyAspMet...AsnAspSerThrLeuAlaValAlaLysIleLeuVa  550
      ||||||::|||                          :::                        |::            ||||||::::
10306  CCAATGACGCCAAAACGCCCTGGATATCCAACGATTACTTTTGACGAAAA  10355       11092  GATGCCATATGACAGGTGATTATACTTTCGGAGTAACGATAACGTACCGCA  11141

289  nThrThrSerValLeuIle.......ArgLysTyrAlaI  300              550  lGluTyrAlaLeuGlnAsnProProGlnLeuThr..........  562
      :::            |||:::::                                         |||||||:::
10356  TACCAAATAGTTACACGTTTGATTTTGGAAAAACCAACAAACGTTACATTA  10405      11142  GGAATATTCCGTGGATGAAGAG...  11185

300  leGlyAspTyrSerLysLeu.......LeuGluGlyAlaThrLeuGln  313      563  .......AspLeuAspPhePheIlePro.  569
      |:|:::||:|||                  |||:||                                  ||||||
10406  TT..GAGTATAAAAACGCCAATGGCTGGAGCGTGCCAACTCTTTTAT  10452         11186  AGCTGGGTCAAAGGAGGAACACCAACTAAAAATTCCATTAACGAAAACAATT  11235

314  LeuThrGly.......AspAsnValAsnSerPheGlnAl  324              570  AsnAsnAsnLysTyrGln.......SerLeuIleGlyThrGl  581
                       :::|||::||                                      :::::::|||                  |::|||
10453  ATAACAGGGACAGCGGAAAGAACCACACAATCGAATAAATAATGAAGGCTCGC  10502     11236  GATCACAGTCGTTTACAAGTCAAAGATTCAACGATTTAATGTCGGCGGATTC  11285

324  aArgValPheSerSerAsnAspIleGlyGluArgIleGluLeuSerAspG  341      581  nTrpHisProGluAspLeuValAspIleIleArgMetGluAspLysLys  597
      :::|||            ||||||                                        |||||            ||||||
```

```
11286 ATGGAAACCAGAAGAG.....AACTTTGTTTCAGCAACAGAGATAAAAACAG 11329
  598 ..GluValIleProValThrHisAsnLeuThrLeuArgLysThrValThr 613
          :::    ||||||  |||||::
11330 GTCAAGACGTTCCTTCGAA.....AAAATCACTGTTTCA 11364

  614 GlyLeuAlaGlyAspArgThrLysAspPheHisGluIleGluLeuVly 630
          :::|||||||::||
11365 GGTCAAGTTGGATAACASCAAAGCACGAGCCGTTTATCCAATTATTACAGTGA 11414

  630 sAsnAsnLysGlnGluLeuSerGlnThrValLysThrAspLysThrA 647
          :::||||:|| |||  ||:::||||
11415 CGAAGGTAAAGAAGAAGAAACAGCCTATGTGACCGTCAAACCCGACCAATCTA 11464

  647 snLeuGluPheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGlu 663
          ::||||||    |||    |||
11465 AGTTGGAGGTCAAAGAT.....ACAACGATTTATGTT......GGTGAT 11502

  664 SerLeuThrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysGl 680
          ||||::||||    ||:::|||:::
11503 TCGTTGGAAA........CCAGA.....GATAATTTCGTTTCAGC 11534

  680 uThrAspSerGluGlyLyrTyrLysVal.........LysValAsnSerGlnG 694
          |||||||||  ||:    |||:|:::||
11535 GACAGACAAAAACAGGTTCCGTTCGAAAAAATTGAAGTTCAGG 11584

  694 luValAlaAsnAlaThrValSerLysThrGlyIleThrSerAspGluThr 710
          |:|||||    |||    |||:|:||
11585 GAACAGTGAAT.....GTTGATAAAATAGGC......GATTATGAA 11619

  711 LeuAlaPheGluAsn 715
          |:::||||:|||
11620 ATTGTCTATAAAAT 11634
```

seq_name: /cgn1_8/gcgdata/geneseq/geneseqn/NA1997.DAT:T68844

seq_documentation_block:
```
ID  T68844 standard; DNA; 5547 BP.
XX
AC  T68844;
XX
DT  29-JAN-1998  (first entry)
XX
DE  Photorhabdus luminescens insect toxin TcdAii.
XX
KW  Insecticide; insect; toxin; pest control; biological control;
KW  Photorhabdus luminescens; TcdA; Southern corn rootworm;
KW  Colorado potato beetle; Western corn rootworm; meal worm;
KW  boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW  cabbage looper; codling moth; corn earworm; European corn borer;
KW  tobacco hornworm; tobacco budworm; Lepidoptera; European corn borer;
KW  Diptera; Dictyoptera; Acarina; Homoptera; ds.
OS  Photorhabdus luminescens strain W-14 (ATCC 55397).
XX
PN  WO9717432-A1.
XX
PD  15-MAY-1997.
XX
PF  06-NOV-1996;  96WO-US18003.
XX
PR  28-AUG-1996;  96US-0705484.
PR  06-NOV-1995;  95US-0007255.
PR  28-FEB-1996;  96US-0608423.
XX
PA  (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI  Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
PI  Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI  Petell J, Roberts JI, Rocheleau TA, Schoonover S;
PI  Strickland JA;
XX
DR  WPI; 1997-281022/25.
DR  P-PSDB; W17900.
XX
PT  Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
PT  can be genetically engineered into insect larvae food and plants for
PT  insect control
XX
PS  Claim 38; Page 201-208; 276pp; English.
XX
CC  This genomic DNA sequence encodes the 209.2 kDa TcdAii insecticidal
CC  toxin protein (see W17899) of Photorhabdus luminescens W-14.
CC  It is derived from a full-length clone (see T68843) obtained from a
CC  cosmid library. A 282.9 kDa protein (see W17899) is proteolytically
CC  processed into TcdAii and TcdAiii (see W18301) components.
CC  Claimed toxins of P. luminescens (see W17871, W17884-89, W17899-900,
CC  W18301-06) can be produced by recombinant DNA methods and applied to,
CC  or genetically engineered into, insect larvae food and plants for
CC  insect control. The Photorhabdus toxins are particularly effective
CC  against Southern corn rootworm, Colorado potato beetle, Western corn
CC  rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet
CC  armyworm, black cutworm, cabbage looper, codling moth, corn earworm,
CC  European corn borer, tobacco hornworm and tobacco budworm
CC  (Lepidoptera), and are also active against insects of the orders
CC  Hymenoptera, Diptera Dictyoptera, Acarina and Homoptera. (All
CC  claimed).
XX
SQ  Sequence 5547 BP; 1710 A; 1219 C; 1115 G; 1503 T; 0 other;
```

alignment_block:
US-09-494-297-2 x T68844

Align seg 1/1   to: T68844   from: 1   to: 5547

alignment_scores:
```
Quality:              154.00        Length:              809
Ratio:                0.424         Gaps:                37
Percent Similarity:   44.870        Percent Identity:    18.294
```

```
  157 AlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGluGl 173
          |||   |||  |||:::| ||:::|| ||
 2386 GCACAACAATTGAATGTCGCCCCCCACAGGCGCGTTTCCGCTTTGGTC...GG 2432

  173 yLeuGluProLeuAsnAlaIleArgValThrGlnGlu...AlaValTrpt 189
          |||::::  ||:||:: ||| |||   |||
 2433 GCTGGATTATATCAATCAAATGAAAGAGACAACGGACCACCGGACCTAAT 2482

  189 yrTyrSerAsp.............AsnAlaProIleSerAsn 198
          :::                   |||:::||||
 2483 AAAACGCGCGCAGGCGGTATTAACCGCCGGGTTGAATTCACAACAGGCTAAT 2532

  199 ProAspGluSerPheLysArgGluSerGluSerAsnLeuValSerThr 214
          |||:::|||:::  |||     |||:::||
 2533 ACAATTACACGCTTTTCTGGATGAATCTCGCAGTGCCGCATTAAGCACCTA 2582

  215 ...SerGlnLeuSerLeuMetArgGlnAlaLeuLysGln... 226
          |||:::|||                  |||:::|
 2583 CTATATCCGTCAAGTCGCCAAGGCAGCGGGCGGCTTATTAAAAGCCGTGATG 2632

  227 ...LeuIleAspProAsnLeuAla... 233
          ||||     |||:::||
 2633 ACTTGTGTATCCAAATACTTACTGATTCGATAATCAGGTTTCTCGGCAATAAAA 2682

  234 ...ThrLysMetProLysGlnValProAspAspPheGlnLeu... 246
          |||:::||  |||    |||:::|
 2683 ACCACCCGGATCGGCGAAGCCATTGCC...AGTRATTCAACTGTACGTCAA 2729

  247 ..........SerIlePheGluSerGluAspLysGly... 255
          :|||::|||      |||:::|
 2730 CCGGGCCATTGGAAAATGTGGAAGAAAAAATGCCAATTCGGGGGGGTTATCAGCC 2779

  256 ...AspLysTyrAsnLysGlyTyrGlnAsnLeu 265
          |||||||  |||      |||:::|
 2780 GCCAAATTCTTTATCGACTCGGCGACAAATACAATAAACGCTACAGCACTGG 2829
```

```
266  LeuSerGlyGlyLeuValProThrLysProGlyThrProGlyAspPr   282
     :::                             ||||||
2830 GCGGGTGTTTCTCAATTAGTTTACTACCGGAAAAACTATATTGATCCGAC 2879

282  oMetProProAsnGlnProGlnThrThrSerValLeuIleArgLysTyrA  299
     ::|||                 |||||::|||:::||::|::   TyrA
2880 CATGCGTATCGGACAAACCAAAAATGATGGAGACCCATTACTGCAATCCGTCA 2929

299  la.......                                          299
     ::
2930 GCCAAACCCAATTAAACGCCGGATACCCTCGAGAGATGCCTTTAnGTCTTAT 2979

300  IleGlyAspTyrSerLysLeuLeuGlyGlyAlaThrLeuGlnLeuThrGl  316
     ::||||||                    ::::
2980 CTGACATCGTTTGAACAAGTGGCTAATCTTAAAGTTATTAGCGGCATATCA 3029

316  yAspAsnAsnSerPheGlnAlaArgValPhe...........   Sers  329
     ||||:::::::||::::                                |||
3030 CGATAAATATTAATAAACGATCAAGGGGTGACCCTATTTTATCGGACTCAGTG 3079

329  erAsnAspIleGlyGlu..........        ArgIleGluGlu  338
     ::                                |||
3080 AAACTGATGCCGGTGAATATTATTGGGCCCAGTGTCGANCACAGTAAAATTC 3129

339  SerAspGlyThrTyrThrLeuThr............       GluLeuAs  349
     ::|||||||                                   ::::
3130 ACGACGGTAAATTCGGGGCTAATGCCTGGAGTGAATGGCATAAAATTGA 3179

349  nSerProAlaGly......    TyrSerIleAlaGluProIleThrPheLysV 364
     :::                   |||              ::|||:||
3180 TTGTCCAATTAACCCTTATAAAAGCACTATCCGTCCAGTGATATATAAA. 3228

364  aIgIuAIaGIyLysValTyrThrIle....IleAspGlyLysGlnIleGlu  379
     :::                           ::||:::::|:::
3229 .......TCCCGCCTGTATCTCGTTGGACAAAAAGGAGATCACC 3270

380  AsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAsnAs  396
     :::                      |||:::|:||:::
3271 AAACAGACAGGAAATAGTAAAGATGGCTATCAAACTGGATTATCG 3320

396  pPheGlu.......                                     398
     ::|||
3321 TTATGAACTAAAATTGGGCAATATCCGCTATGATGGCACTTGGAAATACGC 3370

399  .......            GluPheSerValLeuThrThrGlnAsn     407
                       ::|||         |||
3371 CAATCACCTTTGATGTCAATAAAAAAATATCCGAGCTAAAACTGGAAAA 3420

408  .......       TyrAlaLysPheTyrTyrTyrAlaLysAsn        416
                  :::||||        ::::||:||
3421 AATAGAGCGCCCCGGACTCTATTGTGCCGGGTTATCAAGGTGAAGATACGTT 3470

417  .......              LysAsnG    419
                         |||||:
3471 GCTGGGTGATGTTTTATAACCACAAGAGCACACTAGATAGTTATAAAACCG 3520

419  lySerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProPro  435
     :::||                     :::||||
3521 CTTCAATGCAAGGACTATATATCTTTT..GCTCGATATGGCATCC. 3561

436  AspSerGluAspGlyLysThrMetThrPro...........         446
     |||                 |||||||||
3562 ..........AAAGATATGACCCCAGAACAGAGCAAATGTTTA 3593

446  .......                                            446
3594 TCGGGATAATAGCTATCAACAATTTGATACCAATAAGTCAGAAGAGTGA 3643

446  .......                                            446
3644 ATAACCCGCTATGGTTGGGGGAGATTATTACCTCAGCCATTGGTATATAACGGA 3693
```

```
447  ...AspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAs  462
        :::               |||
3694 GATGACTATGTTTTGGGGAGANTTAATTACCCTCAGCCATGGTATATAACGGAGA 3743

462  pLeuPheLysTyrTyrThrValLysProArgAspThrAspProAspThrPheL 479
     |::              :::               ::||
3744 TATTCCACTATCAATTACAATTACAAAGCCGCATCAAGTGATTTAAAATCTATA 3793

479  euLysHisIleLeuLysSerValIleGluLeuLysGlyTyr.......  Arg  491
     :::||::||:::::::                                     |||
3794 TCTCACCAAAATTAAGAAATTATTCATAAATGGATATGAAGGACAGAAGCGC 3843

492  GluLysGlyGlnAlaIleLeuTyrSerGlyLeuThrGluThrGlnLeuAar  508
     |||:::             :::       ::||
3844 AATCAATGCAATCTGATGAATAAAAATGGCAAACTAGGTGATAAATTTAT 3893

508  gAlaAlaThrGlnLeuAlaIleTyrTyrPheThrHisAspSerAlaGluLeuA 525
     :::||            |||:::::||||
3894 TGTTTATACTTGGGGGTCAATCCAAATAACTCGTCAAATAAGCTCA 3943

525  spLysAspLysLeuLysAspTyrHisLeuGlyPheGlyAspMetAsnAspSer 541
     :::||                   :::||                |||
3944 TGTTTTACCCGTCTCTATCAATATACCGGA..........  AACACCAGT 3981

542  ThrLeuAlaValAlaLysIleIleLeuValGluTyrAlaGlnAspSerAsnPr 558
     |||                        :::||
3982 GGACTCAATCAAGGGAGGACTACTA......  TTCCACGGTGACACCACTTA 4025

558  oProGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLys.....  573
     |||        :::||||
4026 TCCA......TCTAAAGTAGAAGCTTGGATTCTGGAGCAAAACGTTCTC 4069

574  .......     TyrGlnSerLeuIleLeuGlyThrGlnTrp.......     582
                |||||                     :::::
4070 TAACCAACCACAAAATGCCGCCATTGGTGATGATTATGCTACAGAGACTCTG 4119

583  ...HisProGluAspLeuValAspIleIleLeuArgMetGluAspLysLysGl  598
        ||||||||||                    :::||
4120 AATAAACCGGATGAGCTTAAGCAATATATCTTTTAGACTGACAGTAAAGG 4169

598  uValIle..........    ProValThrHisAsnLeuThrLeu....     608
     :::||                ::::||
4170 GACTGCTACTGATGTCTCAGGCCCAGTAGAGATTAATACTGCAATTTCTC 4219

609  ....ArgLysThrValThrGlyLeuIleAlaGlyAspArgThrLysAspPhe  623
         :::||                         :::|||
4220 CAGCAAAAGTTCAGATAAATAGTCAAAGCGGGTGGCAAGGAGCAAAACTTTTT 4269

624  HisPheGlu..........                               626
     :::||
4270 ACCGAGATAAAAGAAGTGTCTCCATTCAGCCAACCACCTAGCTTTGATGAAAT 4319

627  .......             IleGluLeuLysAsnAsnLysGlnGluLeuL  637
                        ::||||:::
4320 GAATTATCAATTTAAATGCCCTTGAAATAGACGGTTCTGGGTCTGAATTTTA 4369

637  euSerGlnThrValLysThrAspLysThrAsnLeuGluPhe...  LysAsp  652
     :::||||||:::::||:::                            ::|||
4370 TTAACAACTCAGCCAGCTAATTGATGTGTTACTTTTTACCGGATGGGAGGAT 4419

653  GlyLys..........                    AlaThrIleAsnLe  659
     |||                                ::::
4420 GGCCCGCAAAACTGGGTTATGAAAAGTTTCAGTATTCCGTTACCCTCAAGGT 4469

659  uLysHisGlyGluSerLeuThrLeuGlnGlyLeuProGlnGlyTyrSerT  676
     :::::||||||||:::                   |||
4470 AAGTACCGATAATGCCCTGACCCTGCACCATAATGAAAATGGTCGCAAT 4519

676  yrLeuValLysGluThrAspSerGluGlyTyrLysValLysAsnSer      692
     |::::||:::::::::||:::                      |||:::
4520 ATATG.............  CAATGGCAATCCTATCGTTACCCCGCCTGAATACT 4557

693  .......    GlnGluValAlaAlaAsnAlaThrValSerLysThrGlyIl  705
4558
```

```
seq_name: /cgn1_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V29930

ID    V29930 standard; DNA; 5547 BP.
XX
AC    V29930;
XX
DT    07-AUG-1998  (first entry)
XX
DE    tcdA gene from the tcd genomic region encoding protein TcdAii.
XX
KW    Photorhabdus luminescens W-14; nematode; symbiotic;
KW    Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;
KW    Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
KW    Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
KW    mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
KW    cabbage looper; coding moth; corn earworm; European corn borer;
KW    Tobacco hornworm; budworm; ds.
XX
OS    Photorhabdus luminescens.
XX
FH    Key       Location/Qualifiers
FT    CDS       1..5547
FT              /*tag= a
FT              /note= "no stop codon given"
XX
PN    WO9808932-A1.
XX
PD    05-MAR-1998.
XX
PF    05-MAY-1997;  97WO-US07657.
XX
PR    06-NOV-1996;  96WO-US18003.
PR    28-AUG-1996;  96US-0705484.
PR    06-NOV-1996;  96US-0743699.
XX
PA    (DOWC ) DOWELANCO.
PA    (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI    Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
PI    Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI    Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI    Strickland JA, Sukhapinda K;
XX
DR    WPI; 1998-1794271/16.
DR    P-PSDB; W56573.
XX
PT    Isolated toxins from Photorhabdus luminescens strains - useful for
PT    control of insect pests

PS    Claim 38;  Pages 231-237; 321pp; English.

CC    The present sequence encodes a protein named TcdAii of the bacterium
CC    Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
CC    nematodes of the Heterorhabditis genus. The bacterium has at least 4
CC    distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
CC    produced from these regions that are associated with insecticidal
CC    activity. The native toxins are secreted proteins. The proteins are
CC    toxic to insects upon exposure and especially when ingested. The
CC    nucleic acid sequence can be used to produce transgenic plants. The
CC    baculoviruses or microbial hosts for toxin production. They can be used
CC    to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC    Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC    Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC    boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC    coding moth, corn earworm, European corn borer or tobacco hornworm
CC    or budworm.
XX
SQ    Sequence 5547 BP; 1711 A; 1223 C; 1112 G; 1501 T; 0 other;
XX
```

```
alignment_scores:
               Quality:     152.00        Length:         809
               Ratio:         0.420       Gaps:            37
Percent Similarity:  44.747   Percent Identity:  18.294

alignment_block:
US-09-494-297-2 x V29930

Align seg 1/1  to: V29930    from: 1  to: 5547

 157 AlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGluGl 173
     GCACAACAATTGAATGTCGCCCACAGGCGCGTTTCCGCGTTGGTC...GG 2432
2386

 173 yLeuGluProLeuAsnAlaIleArgValThrGlnGlu...AlaValArgT 189
     ||||:||       :|||:|||||||||||||||||       ||| |||
2433 GCTCGGATTATATATCAATCAATCAAATGAAAGAGACACCGACCTATGCCCAGTGGG 2482

 189 yrTyrSerAsp...........................AsnAlaProIleSerAsn 198
                                            :::|     |||||||||
2483 AAAACGGCGAGGCGTATTAACCGCCGGGGTTGAATTCACAACAGGCTAAT 2532

 199 ProAspGluSerPheLysArgGluSerGluSerAsnLeuValSerThr. 214
     ::::||||:|||||||||||||||||       ::|||||||||
2533 ACATTACACGCTTTTCTGGATGATGAATCTCCCAGTGCCGCAATTAAGCACCTA 2582

 215 .....SerGlnLeuSerLeuMetArgGlnAlaLeuLeuLysGln....... 226
                ||||||:::||||:::||||       |||||||:::
2583 CTATATCCGTCAAGTCGCCCAAGCGCAGCGCGCGCTATTAAAAGCCGTGATG 2632

 227 ..............LeuIleAspProAsnLeuAla............... 233
                  |||||||| |||||:::
2583 ..............LeuIleAspProAsnLeuAla............... 233

 234 ..ThrLysMetProLysGlnValProAspPhePheGlnLeu......... 246
     |||:|||||           :::|||||||||||
2683 ACCACCCGGATCGCGCAAGCCATTGCC...AGTATTCAACTGTACGTCAA 2729

 247 ...........SerIlePheGluSerGluAspLysGly............ 255
                |||||||:::||||       |||||||
2730 CCCGGGCCATTGGAAAATGTGGAAGAAAAATCCCAATTCGGCGGGTTATCAGCC 2779

 256 ..........AspLysTyrAsnLysGlyTyrArgGluAsnLeu 265
           :::|||||||||||||||||||       :::|||
2780 GCCAATTCTTTATCGACTGGGACAAATACAATAAACGGCTACACGCACTTGG 2829

 266 LeuSerGlyGlyLeuValProThrLysProProThrProGlyAspProPr 282
     :::|||       :::|||||||||||       |||||||
2830 GCGGGTGTTTTCTCAAAAGTTAGTTTACTACCCGGAAAAACTATATATGGAATCCGAC 2879

 282 oMetProProAsnGlnProGlnThrThrSerValLeuIleArgLysTyrA 299
     :::|||:||||||||||||||||||||||||       :::||| :::
2880 CATGCCGTATCGGACAAACCAAAAAATGATGACGCATTACTGCCAATCCGTCA 2929

 299 la.................................... 299
     :::
2930 GCCAAAAGCCAATTAAAACGCCGATACCGTCGAAGATGCCTTTATGTCTTAT 2979

 300 IleGlyAspTyrSerLysLeuLeuGluGlyAlaThrLeuGlnLeuThrGl 316
     :::::|||::::|||::::::::|||::::::||||::::::
```

```
4678 CCTCCCTATAACCTATCAACTCATGGT 4704
     ::::||||::|:
4642 ..........................CAGTTAGGCAAAGGTTTCTATCCTACGTTCGTGATA 4677

4558 CTATTTGCCGCGCCAGTTGGTTGCACGCGCCACC.........ACCGGAAT 4598
     ||||||||:::||:|||       :::|||||||
4599 CGATACAATTCTGAGTATGGAAAACTCAGAAATTCAGGAACCG 4641

4642 roThrGlyValAspGlnLysIleAsnGlyTyrLeuAlaLeuIleValIle 738
     :::||       |||:::|||       ||| ||||||
722 roThrGlyValAspGlnLysIleAsnGlyTyrLeuAlaLeuIleValIle 738

705 eThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProValAlaP 722
     ||:||:|       |||:::|||||||||||
4599 CGATACAATTCTGAGTATGGAAAACTCAGAAATTCAGGAACCG 4641

739 AlaGlyIleSerLeuGlyIleTrpGly 747
```

seq_name: /cgn1_8/ycgdata/geneseq/geneseqn/NA1997.DAT:T68843

seq_documentation_block:
```
ID   T68843 standard; DNA; 7551 BP.
XX
AC   T68843;
XX
DT   29-JAN-1998  (first entry)
XX
DE   Photorhabdus luminescens insect toxin TcdA DNA.
XX
KW   Insecticide; insect; toxin; pest control; biological control;
KW   Photorhabdus luminescens; TcdA; Southern corn rootworm;
KW   Colorado potato beetle; Western corn rootworm; meal worm;
KW   boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW   cabbage looper; coding moth; corn earworm; European corn borer;
KW   tobacco hornworm; tobacco budworm; Lepidoptera; European corn borer;
KW   Diptera, Dictyoptera; Acarina; Homoptera; ds.
XX
OS   Photorhabdus luminescens strain W-14 (ATCC 55397).
XX
PN   WO9717432-A1.
XX
PD   15-MAY-1997.
XX
PF   06-NOV-1996;  96WO-US18003.
XX
PR   28-AUG-1996;  96US-0705484.
PR   06-NOV-1995;  95US-0007255.
PR   28-FEB-1996;  96US-0608423.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
PI   Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI   Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI   Strickland JA;
XX
PT   Claim 38; Page 185-194; 276pp; English.
XX
PS   Claim 38; Page 185-194; 276pp; English.
XX
CC   This genomic DNA sequence encodes the 282.9 kDa TcdA insecticidal
CC   toxin protein (see W17899) of Photorhabdus luminescens W-14.
CC   It was isolated by screening a W-14 genomic cosmid library using a
CC   radiolabeled probe. The 282.9 kDa protein is proteolytically
CC   processed into TcdAii and TcdAiii components (see W17900 and W18301).
CC   Claimed toxins of P. luminescens (see W17871, W17884-89, W17899-900,
CC   W18301-06) can be produced by recombinant DNA methods and applied to,
CC   or genetically engineered into, insect larvae food and plants for
CC   insect control. The Photorhabdus toxins are particularly effective
CC   against Southern corn rootworm, Colorado potato beetle, Western corn
CC   rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet
CC   armyworm, black cutworm, cabbage looper, coding moth, corn earworm,
CC   European corn borer, tobacco hornworm and tobacco budworm
CC   (Lepidoptera), and are also active against insects of the orders
CC   Hymenoptera, Diptera Dictyoptera, Acarina and Homoptera. (All
CC   claimed).
XX
SQ   Sequence 7551 BP; 2259 A; 1739 C; 1595 G; 1958 T; 0 other;
```

alignment_scores:
```
Quality:         152.00        Length:         809
Ratio:           0.415         Gaps:            38
Percent Similarity:  45.241    Percent Identity:  18.418
```

alignment_block:

```
US-09-494-297-2 x T68843

Align seq 1/1   to: T68843    from: 1   to: 7551

157  AlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGluGl 173
     |||    |||||||||           |||:||::|||::|||::||   ||
2650 GCACAACAATTGAATGTCGCCCCCAACAGGCGCGTTTCCGCGTTTGGTC...GG 2696

173  yLeuGluProLeuAsnAlaIleArgValThrGlnGlu...AlaValTrpI 189
     |||||||       :::                   |||    :||||   :
2697 GCTGGATTTATTATTCAATCAATGAAAGAGACACCCGACCATGCCCAGTGG 2746

189  yrTyrSerAsp.............................AsnAlaProIleSerAsn 198
     |:::|                                   ||    ||:||
2747 AAAACGCCGCAGGCGTATTAAACCGCCGGGTTGAATTCACAACAGGCTAAT 2796

199  ProAspGluSerPheLysArgGluSerGluSerAsnLeuValSerThr.. 214
     :::::|||||||      :|||                     :||||||||
2797 ACATTACACGCTTTTCTGGATGAATCTCGCCAGTGCCGGCATTAGCACCTA 2846

215  .........SerGlnLeuSerLeuMetArgGlnAlaLeuLeuLysGln..   226
     |||||:                       |||||||::||
2847 CTAATACCGTCAAGTCGCCAAGGCGCGGCGTATTAAAAAGCCGTGATG 2896

227  ..............LeuIleAspProAsnLeuAla...............  233
     ||||||||||          :::|||||:
2897 ACTTGTATCAAATACTTACTGATTGATATATCAGGTTTCTGCGGCAATAAA 2946

234  ..ThrLysMetProLysGlnValProAspAspPheGluLeu........  246
     |||:::||||                     :|||||||
2947 ACCACCCGGATCGCCGAAGCCCATTGCC...AGTATTCAACTGTACGTCAA 2993

247  .........SerIlePheGluSerGluAspLysGly............   255
     |||||||                 ||||||:||
2994 CCGGGCCATTGGAAAAATGTCGAAGAAAAATGCCAATTCGGGGGGTTATCAGCC 3043

256  ............AspLysTyrAsnLysGlyTyrGlnAsnLeu.......   265
     ||||||||          :::||||||:|||
3044 GCCAATTCTTTATCGACTGGACAAAATACAATAAACGCTACACGACACTTGG 3093

266  LeuSerGlyValProThrLysProProThrProGlyAspProPro     282
     |||                     ::|            |||||||
3094 GCGGCTGGTTTTCTCAATTAGTTTACTACCCGGAAAAACTATATTGATCCGAC 3143

282  oMetProProAsnGlnProGlnThrThrSerValLeuIleLeuArgLysTyrA 299
     ||||:::                     |||||||
3144 CATGCGTATCGGACAAAACCAAAAATGATGGACGCATTACTCCAATCCGTCA 3193

299  la...............................................  299
     :::
3194 GCCAAAGCCAATAAACGCGCGATACCGTCGAAGATGCCTTTAGTCTTTAT 3243

300  IleGlyAspTyrSerLysLeuGluGlyAlaThrLeuGlnLeuThrGl   316
     |:::||||||:::|                           :|||||
3244 CTGACATCGTTTGAACAAGTGGCTAAATCTTAAAGTTATTAGCGGATAATCA 3293

316  yAspAsnValAsnSerPheGlnAlaLeuArgValPhe............Ser 329
     |:||||||:||||                               |||
3294 CGATAATAATTAATAACGATCAAAGGGCTGACCTATTTATTCGGACTCAGTG 3343

329  erAsnAspIleGlyGlu...........................ArgIleGluLeu   338
     ||||||:|||                                  :||
3344 AAACTGATGCCGGTGAATATTATTGGGCGCAGTGTCGATCACAGTAAATTC 3393

339  SerAspGlyThrTyrThrLeuThr...................GluLeuAs   349
     :::|||||||                             ::||||
3394 AACGACGGTAAAATTCGCGCGCTAAGCCTGAGTGAATGGCATAAAAATTGA 3443

349  nSerProAlaGly........TyrSerIleAlaGluProIleThrPheLysV 364
     |||  |||              :::|||:|||::
3444 TTGTTCCAATTAACCCTTATAAAAACCACTATCCGTCCAGTGATATAATAAA 3492

364  alGluAlaGlyLysLysValTyrThrIle...IleAspGlyLysGlnIleGlu 379
```

```
3493 ........................:::::|||..::::
            TCCCCCTGTATCCTCTGGTTGGAACAAAAGAGGATCACC 3534

3493 ........................:::::|||..::::

380 AsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAsnAs 396
           ::::.....:::|||.:...|||
     AAACAGACAGGAAATAGTAAAGATGGCTATCAAACTGAAACGGATTATCG 3584

3535 AAACAGACAGGAAATAGTAAAGATGGCTATCAAACTGAAACGGATTATCG 3584

396 pPheGlu................................ 398
         ::::|||
     TTATGAACTAAAAATTGGGGCATATCCGCTATGGACTTGGAATACGC 3634

3585 TTATGAACTAAAAATTGGGGCATATCCGCTATGGACTTGGAATACGC 3634

399 ...............GluPheSerValLeuThrThrGlnAsn 407
     CAATCACCTTTGATGTCAATAAAAAATATCCGAGCTAAAACTGGAAAAA 3684

3635 CAATCACCTTTGATGTCAATAAAAAATATCCGAGCTAAAACTGGAAAAA 3684

408 ..............TyrAlaLysPheTyrTyrAlaLysAsn. 416
     AATAGAACGCGCCCGGACTCTATTGTGCCGGTTATCAAGGTGAAGATACGTT 3734

3685 AATAGAACGCGCCCGGACTCTATTGTGCCGGTTATCAAGGTGAAGATACGTT 3734

417 ............................LysAsnG 419
     GCTGGTGATGTTTTATAACCAACAAGACACTAGAGTTATATAAAAACG 3784

3735 GCTGGTGATGTTTTATAACCAACAAGACACTAGAGTTATATAAAAACG 3784

419 lySerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProPro 435
        ::||||.::|||     ||||||||:::
     CTTCAAATGCAAAGACTATATATACTTTT...GCTGATATGGCATCC. 3825

3785 CTTCAAATGCAAAGACTATATATACTTTT...GCTGATATGGCATCC. 3825

436 AspSerGluAspGlyGlyLysThrMetThrPro.............. 446
         |||     ||||||||
     AAAGATATGACCCCAGAACAGAGCAATGTTTA 3857

3826 AAAGATATGACCCCAGAACAGAGCAATGTTTA 3857

446 ....................................... 446
     TCGGGATAATAGCTATCAACAATTTGAGACCAATAAGTCAGAAGAGTGA 3907

3858 TCGGGATAATAGCTATCAACAATTTGAGACCAATAAGTCAGAAGAGTGA 3907

446 ....................................... 446
     ATAACCGCTATGCGAGAGGATTATGAGAATTCCTTCCTCGGTAAGTAGCCGT 3957

3908 ATAACCGCTATGCGAGAGGATTATGAGAATTCCTTCCTCGGTAAGTAGCCGT 3957

447 ........AspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAs 462
        |||:.|     ||||||
     AAAGACTATGGTTGGGGAGATTATTACCTCAGCATGGTATATAACGGAGA 4007

3958 AAAGACTATGGTTGGGGAGATTATTACCTCAGCATGGTATATAACGGAGA 4007

462 pLeuPheLysTyrThrValLysProArgAspThrProAspThrProPheL 479
     |||     ::|||     ::::
     TATTCCAACTATCAATTACAAAGCCGCATCAAGTGATTTAAAAATCTATA 4057

4008 TATTCCAACTATCAATTACAAAGCCGCATCAAGTGATTTAAAAATCTATA 4057

479 euLysHisIleLeuLysSerValIleLeuGluLysGlyTyr........ 491
     ::||||||     |||     |||
     TCTCACAAAAAATTAAGAATTAATTCATAATGGATATGAAGGACGAAAGCGC 4107

4058 TCTCACAAAAAATTAAGAATTAATTCATAATGGATATGAAGGACGAAAGCGC 4107
                                                    Arg 491

492 GluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGluThrGlnLeuAr 508
     ::|||     ::|||
     AATCAAATGCAATCTGATGAATAAAATATGGCAAACTAGTGATAAATTTAT 4157

4108 AATCAAATGCAATCTGATGAATAAAATATGGCAAACTAGTGATAAATTTAT 4157
                                              LeuLysAspSer 508

508 gAlaAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuA 525
     .:::||||||::||.:.::|||     ||.::|||
     TGTTTATACTGCTTGGGGTCAATCCAAATAACTCGTCAAATAAGCTCA 4207

4158 TGTTTATACTGCTTGGGGTCAATCCAAATAACTCGTCAAATAAGCTCA 4207
                                          HisAspSerAlaGluLeuA 525

525 spLysAspLysLysAspTyrHisGlyPheGlyAspMetAsnAspSer 541
     ::::|||     |||
     GGAACTCAATCAAGGAGACTACTA...........ACACCAGT 4245

4208 GGAACTCAATCAAGGAGACTACTA...........ACACCAGT 4245
                                       AspMetAsnAspSer 541

542 ThrLeuAlaValAlaLysIleLeuValGluTyrAlaGlnAspSerAsnPr 558
     |||     |||
     TTCCACCGTGACCACACTTA 4289

4246 GGACTCAATCAAGGAGACTACTA.......TTCCACCGTGACCACACTTA 4289
                                          AspSerAsnPr 558

558 oProGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLys.... 573
     :.|||::::::|||||||||:::
     TCTAAAGTAGAAGCTTGGATTCCTGGAGCAAAAAGTTCTC 4333

4290 TCCA..........TCTAAAGTAGAAGCTTGGATTCCTGGAGCAAAAAGTTCTC 4333

574 ...........TyrGlnSerLeuIleGlyThrGlnTrp......... 582
     ::::|||     ||||||
                                          582
```

4334 TAACCAACCAAAAATGCGGCACCATTGGTTGATGATTATATGCTACAGAGACTCTCTG 4383
         :::::|||:::
583 ...HisProGluAspLeuValAspIleIleArgMetGluAspLysGlyl 598
         |||::|||:||||| ||| |||.::||||||:::|||
4384 AATAAACCGGATAGATCTTAAGCAAATATATCTTTAGACTGACAGTAAAGG 4433

598 uValIle................ProValThrHisAsnLeuThrLeuArgl 610
            ||||| ::::::
4434 GACTGCTACTGATGTCTCAGGCCCAGTAGACAGATTAATACTGCAAТTTCTC 4483

610 ysThr............ValThrGlyLeuAlaGlyAspArgThr... 620
        ::: ||| ::::::|||
4484 CAGCAAAAGTTCAGATAAТAAGTCAAAACGGTTGGCAAGGAGCAAAACTTTT 4533

621 ....LysAspPheHisPheGlu.......... 626
            ||||||
4534 ACCGGAGATAAAAGATGTCTCCATTCAGCCAТCACCTAGCCTTTGATGAAAT 4583

627 ....IleGluLeuLysAsnAsnLysGlnGluLeuL 637
            ::::::
4584 GAATTATCAATTTAAATGCCCTTGAAATAGACAGGTTCTGGTTCTGAAТTTTA 4633

637 euSerGlnThrValLysThrValAspLysThrAsnLeuGluPhe...LysAsp 652
     :::::::|||:::|||||||::||| :::|||
4634 TTAACAACTCAGCCAGTATTGAGTTGTTACTTTTACCCGCATTTGGCGGAGGAT 4683

653 GlyLys...............AlaThrIleAsnLe 659
     |||::: |||::::::
4684 GGCCCAAAACTGGGTTGTTATGAAAGTTTCAGTAТTCCTGTTACCCTCAAGGT 4733

659 uLysHisGlyGluSerLeuThrLeuGlnIleGlyLeuProGluGlyTyrSerT 676
            ::::|||::::::||||||
4734 AAGTACCGATAATGCCCTGACCCTGCCACCAATAAТGAAAATGGTGCCAAT 4783

676 yrLeuValLysGluThrAspSerGluGluGlyTyrTyrLysValLysValAsnSer 692
     |||:::   ::::|||:::
4784 ATATG.............CAAТGGCAAТCCTAТCGTACCCGCCCTGAAТTACT 4821

693 ....GlnGluValAlaAlaAsnAlaThrV, AlSerLysThrGlyIl 705
            |||||||||:::|||||||::||||||
4822 CTATTTGCCCGCCCAGTTGGGTTGCACGCCCACC...ACCGGAAТ 4862

705 eThrSerAspGluThrLeuAlaPheLeuAsnAsnLysGluProValValP 722
     :::   ::|||::||||||
4863 CGATACAATCTGAGTATGGAACACCAGAAТAAТTCAGGAACCG........ 4905

722 roThrGlyValAspGlnLysIleAsnGlyTyrTyrLeuAlaLeuIleValIle 738
     :::||||||::::|||::|||:::||||||
4906 ...........CAGTTAGGCAAAGGTTTCTAТGCTACGTTCGTGATA 4941

739 AlaGlyIleSerLeuGlyIleTrpGly 747
     ::|||:: |||
4942 CCTCCCTTATAACCTAТCAACTCATGGT 4968

OS   Photorhabdus luminescens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..7551
FT   /*tag= a
XX
PN   WO9808932-A1.
XX
PD   05-MAR-1998.
XX
PF   05-MAY-1997; 97WO-US07657.
XX
PR   06-NOV-1996; 96WO-US18003.
PR   28-AUG-1996; 96US-0705484.
PR   06-NOV-1996; 96US-0743699.
XX
PI   Strickland JA, Sukhapinda K;
PI   Petell J, Roberts JL, Rocheleau TA, Schoonover S;
PI   Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
PI   Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;
XX
PA   (DOWC) DOWELANCO.
PA   (WISC) WISCONSIN ALUMNI RES FOUND.
XX
DR   P-PSDB; W56572.
DR   WPI; 1998-179427/16.
XX
PT   Isolated toxins from Photorhabdus luminescens strains - useful for
PT   control of insect pests
XX
PS   Claim 38; Pages 215-224; 321pp; English.
XX
CC   The present sequence encodes a protein named TcdA of the bacterium
CC   Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
CC   nematodes of the Heterorhabditis genus. The bacterium has at least 4
CC   distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
CC   produced from these regions that are associated with insecticidal
CC   activity. The native toxins are secreted proteins. The proteins are
CC   toxic to insects upon exposure and especially when ingested. The
CC   nucleic acid sequence can be used to produce transgenic plants,
CC   baculoviruses or microbial hosts for toxin production. They can be used
CC   to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC   Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC   Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC   boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC   codling moth, corn earworm, European corn borer or tobacco hornworm
CC   or budworm.
XX
SQ   Sequence 7551 BP; 2256 A; 1737 C; 1600 G; 1958 T; 0 other;

alignment_scores:
          Quality: 152.00
            Ratio: 0.420
Percent Similarity: 44.747   Percent Identity: 18.294

alignment_block:
US-09-494-297-2 x V29928

Align seg 1/1 to: V29928 from: 1 to 7551
US-09-494-297-2   to: V29928   ..
                                       Length: 809
                                         Gaps: 37

157 AlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyIleMetGluGl 173
    ||| ||||:::::|||||||||::::::::::| ||
2650 GCACAACAATTGAAATGTCGCCCCACAGGCGCGTTTCCCGCTTTGGTC...GG 2696

173 yLeuGluProLeuAsnAlaIleArgValThrGlnGlu.....AlaValTrp 189
    |||:::||| ||| ||| |||
2697 GCTGGATTATTATTCAAATCAATGAAAGAGACACCGACCTATGCCAGTGGG 2746

189 YrTyrSerAsp......................AsnAlaProIleSerAsn 198
      :::||| ::::||| ::|||
2747 AAAACGCGGCAGGCGGTGATTAACCGCGCGGGGTTGAATTCACAACAGGCTAAT 2796

199 ProAspGluSerPheLysArgGluSerGluSerAsnLeuValSerThr.. 214
    ::::|||:::||||||| |||||::||| ::::|||||||||
2797 ACATTACACGCCTTTTCTGGATGAGAATCTCCGACGTGCCAGTTGCAATTAAGCACCTA 2846

215 ....SerGlnIleuSerIleuMetArgGlnAlaLeuLysGln........ 226
          ||||:::||| ||||||| :::|||||||||
2847 CTATATCCGTCAAGTCGCCAAGGCAGCCGCGCGGTCGTTATTAAAAGCCGTGATG 2896

227 ....LeuIleAspProAsnLeuAla................ 233
                 ||||||||| |||
2897 ACTTGTATCACAATACTTACTGCATTGATAATCCAGGTTTTCTCCGGGCAATAAAA 2946

234 ...ThrLysMetProLysGlnValProAspApsPheGlnLeu...... 246
       ||||||||||||||||||||||::::|||
2947 ACCACCCGGATCGCCGAAGGCCATTGCC...AGTATTCAACTGTAGCGTCAA 2993

247 ....SerIlePheGluSerGluAspLysGly............ 255
             :::||| |||||||::::::|||
2994 CCGGGCCATTGGAAAATGTGGAAGAAAAATGCCCAATTCGGCGGGGTTATCAGCC 3043

256 ....AspLysTyrAsnLysGlyTyrGlnAsnLeu.......... 265
        |||||||||||||||||||| ::::|||
3044 GCCAATTCTTTATCGACTGGGACAAAATACAATAAAACGCTACAGCACTTGG 3093

266 LeuSerGlyLeuValProThrLysProProThrProGlyAspProPr.... 282
    |||||| :::|||::::::||||| :::::|||||||
3094 GCGGGTGTTTCTCAATTAGTTTACTACCCGGAAAACTATATTGATCCGAC 3143

282 oMetProProAsnGlnProGlnThrThrSerValLeuIleLeuArgLysTyrA 299
    :::||| |||||||::::::|||:::::|||||||||::::|||;
3144 CATGCGTACGGACAAACCAAAAATGATGGACGCATTACTGCAATCCGTCA 3193

299 la....................................... 299
    |||
3194 GCCAAAGCCAATTAAACGCCGATACCGTGCAAGATGCCTTTATGTCTTAT 3243

300 IleGlyAspTyrSerLysLeuGlyGlyAlaThrLeuGlnLeuThrGl 316
    :::||||||||||||||||||||||| |||||||||||||
3244 CTGACATCGTTTGAACAAGTGGCTAATCTTAAAGTTATTAAGCGCATATCA 3293

316 yAspAsnValAsnSerPheGlnAlaArgValPhe................ 329
    |||||||::::::||| ::::::||| Ser5 329
3294 CGATAATAATAACGATCAAGGGCTGACCTATTTATCGGACTCAGTG 3343

329 erAsnAsnAspIleGlyGlu...........................ArgIleGluLeu 338
    :::|||::::|||||||:::   ::::|||
3344 AAACTGATCGCCGGTGAATATTATTGGCGCAGTGTCGATCACAGTAAATTC 3393

339 SerAspGlyThrThrTyrThrLeuThr....................GluLeuAs 349
    ::::|||||||::::::||| ::::|||
3394 AACGACGGTAAAAATTCCGGCCTAANGCCTGGAGTGAATGGCATAAAATTGA 3443

349 nSerProAlaGly......TyrSerIleAlaGluProIleThrPheLysV 364
    :::||| ||| ::::||| :::::|||
3444 TTGTCCAATTAACCCTTATANAAAGCACTATCCGTCCTTCAGTGATAATATAA 3492

364 alGluAlaGlyLysValTyrThrIle...IleAspGlyGlyLysGluIleGlu 379
    ::::::||||||||||||||||||| |||||||::::::|||
3493 ........TCCCGGCCTGTATCTGCTCTGGTTGGACAAAAGGAGATCACC 3534

380 AsnProAsnLysGluIleValGluProTyrSerValGlyAlaTyrAsnAs 396
    ::::|||||||::::::||| :::::|||::::::|||
3535 AAACAGACAGGAAATAGTAAAGATGGCTATCAAACTGAAAACGGATTATCG 3584

396 pPheGlu.. 398
    :::|||
3585 TTATGAAACTAAAAATTGGGCCATATCCGCTTATGATGGCACTGGAATACGC 3634

399 ....GluPheSerValLeuThrThrGlnAsn............ 407
               :::||| ||| ||||
3635 CAATCACCTTTGATGTCAATAAAAAAAATCCGAGCTAAAACTGGAAAAA 3684

408 ....TyrAlaLysPheTyrTyrAlaLysAsn...... 416

3685 AATAGAGCGCCCGGACTCTATTGTGCCGGTTATCAAGGTGAGATACGTT 3734
     ||| :::::::
     ::::::   LysAsnG 419
417 ....

3735 GCTGGTGATGTTTTATAACCACACAAGACACACTAGATAGTTATAAAAACG 3784
     ||:| |:|::||| |||||||:|||||||| ||
419 LysSerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProPro 435

3785 CTTCAATGCAAGGACTATATATCTTT..GCTGATATGGCATCC. 3825
     |||| |:||||:||| |||
436 AspSerGluAspGlyGlyLysThrMetThrPro............. 446

3826 ..............AAAGATATGACCCCAGAACAGAGACAAATGTTTA 3857
                  ||||||||||
446 446

3858 TCGGGATAATAGCTATCAACAATTTGATACCAATAAATGTCAGAAGAGTGA 3907
     || ||:
446 ....

3908 ATAACCGCTATGCAGAGGATTATGAGATTCCTTCCCTCGGTAAGTAGCCGT 3957
     || ||:||
446 ....GlnGluValAlaAsnAlaThrValSerLysThrGlyIll 705

447 ...AspPheThrThrGlyGlyValValLysThrYrThrHisIleAlaGlyTyrArgAs 462
     ||:||| |:
3958 AAAGACTATAGGTTGGGAGGAGATTATTACCTCAGCATGGTATATAACGGAGA 4007

462 pLeuPheLysTyrThrValLysProArgAspThrAspProAspThrPheL 479
     ||| |:|| || |||||:||||||||||
4008 TAATTCCAACTATCATTTACAAAGCCGCATCAAGTGATTTAAAAATCTATA 4057

479 euLysHisIleLysLysValIleLeuGluLysGlyTyr............. 491
     |||| |||
4058 TCTCACCAAAATTAAGAATTATTCATAATGGTATATGAAGGACAGAAGGC 4107 Arg 491

492 GluLysGlyGlnAlaIleLeuThrTyrSerGlyLeuThrGlnLeuAr 508
     ||| ||| |||
4108 AATCAAATGCAAATCTGATGAATAAAAATATGGCAAACTAGTGATAAATTTAT 4157

508 gAlaAlaThrGlnLeuAlaIleTyrThrPheThrAspSerAlaGluLeuA 525
     ||| ||| ::||| :::||
4158 TGTTTTACTACTTGGGGTGTCAAATCAAATAACTCGTCAAAATAAGCTCA 4207

525 spLysAspLysLeuLysAspThrYrHisGlyPheGlyAspMetAsnAspSer 541
     ||:||:|| |||
4208 TGTTTTACCCCGTCTATCAATATAGCGGA.........AACACCAGT 4245

542 ThrLeuAlaValAlaLysIleLeuValGluTyrAlaGlnAsnAspAsnPr 558
     :::|||||| :::||||||
4246 GGACTCAATCAAGGAGACTACTA...........TTCCACCGTGACACACTTA 4289

558 oProGlnLeuThrHisAspLeuAspPhePheIleProAsnAsnAsnLys... 573
     ||| |||
4290 TCCA.........TCTAAAGTAGAAGCTTGGATTCCTGGACAAAAACGTTCTC 4333

574 ......TyrGlnSerLeuIleGlyLeuThrGlnTrp................. 582
           |||||| |:||||
4334 TAACCAACCAAAATGCCCCCATTGGTGATGATTATGCTACAGAGACTCTCTG 4383

583 ...HisProGluAspLeuValAspIleIleArgMetGluAspLysLysGl 598
     ||:||||| |||:::||
4384 AATAAAACCGGATGATGTCTTAAGCAAATATATCTTTATGACTGACAGTAAAGG 4433

598 uValIle............ProValThrHisAsnLeuThrLeu...... 608
     ||:|||| |||||||
4434 GACTCGCTACTGAATGTCTCAGGCCCCAGTAGAGAATTAATACTGCAATTCTC 4483

609 .....ArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPhe 623
          ||:||| |||
4484 CAGCAAAAGTTCAGATAATAGTCAAAGCGGGTGGCAAGGAGCAAACTTTT 4533

624 HisPheGlu.. 626
     ::
     ::::::

4534 ACCGCAGATAAAGAGATGTCTCCATTCAGCCACCATCACCTAGCTTTGATGAAAT 4583
                ::|||::LeuGluLeuLysAsnAsnLysGlnGluLeuL 637
627 ....IleGluLeuLysAsnAsnLysGlnGluLeuL 637

4584 GAATTATCAATTTAATGCCCTTTGAAAATAGACGGTTCTGGGTCTCGAATTTA 4633
     :::|||::||||
637 euSerGlnThrValLysThrAspLysThrAsnLeuGluPhe...LysAsp 652

4634 TTAACAACTTCAGCCACAGTATTTTGATGTTTTACTTTTACCGCATTTGCGGAGGAT 4683
     ||| ||| ||||
653 GlyLys..................................AlaThrIleAsnLe 659

4684 GGCCGCAAACTGGGTTATGAAAAGTTTCAGTATTCCCTGTTAACCTCAAGGT 4733
     |||:::||| ::|||||::||
659 uLysHisGlyGluSerLeuThrLeuGlnIleGlyLeuProGlnHisGlyTyrSerT 676

4734 AAGTACCGATAAATGCCCTGACCCTGCCACCATAAATGAAAAATGGTGGCAAT 4783
     :::|||| |::|||
676 yrLeuValLysGluThrAspSerGluGlyTyrLysValLysValAsnSer 692

4784 ATATG..........CAATGGCAAATCCTATCGTACCGCCCTGAATACT 4821
     :: :::|||::
693 ....GlnGluValAlaAsnAlaThrValSerLysThrGlyIll 705

4822 CTATTTGCCCGCCCAGTTGGGTTGCACGCGCCACC..........ACCGGAAT 4862
     ||| |||||:::|||||||| |||||||
705 eThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProValValP 722

4863 CGATACAAATTCTGAGTATGCAAAACTCAGAATATTCAGGAACCG........ 4905
     :::   ::|||||||
722 roThrGlyValAspGlnLysIleAsnGlyTyrLeuAlaLeuIleValIle 738

4906 ..............CAGTTAGGCAAAAGGTTTCTATGCTACGTTCGTACGGTGATA 4941
     |||||||||
739 AlaGlyIleSerLeuGlyIleTrpGly 747
         |||
4942 CCTCCCTTATAACCTTATCAACTCATGGT 4968
     :::|||::::

```
PT   Isolated DNA that encodes group A Streptococci fibrinogen and
PT   fibronectin binding proteins, useful for the prevention, diagnosis
PT   and treatment of Streptococcal infections
XX
PS   Claim 2; Fig 3A-B; 17pp; English.
XX
CC   This invention describes the isolation of a novel group A Streptococci
CC   fibrinogen and fibronectin binding protein (SFFBP-12) and its encoding
CC   nucleic acid which is involved in adhesion of the bacteria to the host
CC   cell, and may initiate the infection process). The products of the
CC   invention are useful in the preparation of compositions and tests for
CC   the diagnosis, treatment and prevention of streptococcal infection.
CC   SFFBP-12 nucleic acid may be used as a hybridization probe to isolate
CC   corresponding genes from other species or in biological samples by, for
CC   example, Northern/southern Blotting. Additionally, the nucleic acid may
CC   also be transfected into host cells and used to recombinantly produce
CC   SFFBP-12 proteins in fermentation cultures. The SFFBP-12 protein may be
CC   used as an antigen in the preparation of vaccines to stimulate a hosts
CC   immune system against Streptococcal infection. Group A Streptococci (e.g.
CC   Streptococcus pyogenes) are the etiologic agents for a range of
CC   suppurative infections (e.g. pharyngitis, impetigo and necrotizing
CC   fasciitis), systemic diseases (e.g. scarlet fever, toxic-shock syndrome)
CC   and may lead to serious sequelae (e.g. rheumatic fever and
CC   glomerulonephritis).
XX
SQ   Sequence 3531 BP; 1231 A; 614 C; 768 G; 918 T; 0 other;

alignment_scores:
       Quality:          150.00
       Ratio:             0.367
       Percent Similarity:  42.516    Percent Identity:  19.854

alignment_block:
US-09-494-297-2 x X77451

Align seg 1/1  to: X77451   from: 1  to: 3531

                                               Length:   962
                                               Gaps:      51

 21 LysAsnSerLysArgPheThrValThrLeuVal......GlyValPheIle 35
    |||:||::||||:|||:|||:||||||||:::::      :|||||||
529 AAAACCACCGGACTTGGACAGTGACTGTTATGAGAACGGCTATACCAA 578

 35 uMetIle.........PheAlaLeuValThrSerMetValGlyAlaIleAla 48
    :::::           ||| ||||:||:|||::|:||::||||:|||
579 GTTGGTTGGAAAATCCCTATAAATGGGGAAATCATCAGTAAAGCAGGGTCAA 628

 48 ysThrValPheGlyLeuValGluSerSerThrProAsnAlaIleAsnPro 64
    |||||     |||||::|||||||||:||||||||||:||||
629 AAGATGTT.......AGTAGTTCTTTTACAGTTGGAAAATCCC 663

 65 AspSerSerSerGluTyrArgTrp.................TyrGlyTy 75
    |||                       |||
664 AAAATGTCAGTTGTTTCTAAATATGGGAAAACAGAGGTTAGTAGTGGCGC 713

 75 rGluSerTyrValArgGlyHisProTyrTyr.......... 85
    ::|:||     |||||:|:||||||| |||
714 AGCGGATTTCTACCGCAACCATGCGCCGCCTATTTAAAATGTCTTTTGAGT 763

 86 .LysGln................PheArg 89
    |||||||                |||
764 TGAAACAAAAGGATAAAAATCTGAAACAATCAACCACCAGTGATACCTTGTG 813

 90 ValAlaHisAspLeuArgValAsnLeuGluGlySerArgSerTyrArgGlnVa 106
    ||| |||||||||||||||:|||||||||||:|::|
814 TTACAGCTGGATAAGAACGTTCTCAATCCTAAAGGT 846

106 lTyrCysPheAsnLeuLeuLysLysAlaPro................ 117
    :::|||                       |||
847 ...........ATCAGTCAAGTATCCCTAAAATCATTACGACAGTG 883

117 euGlySerAspSerSerValLysLysTrpTyrLysLysHisAspGlyIle 133
    ::|||| |   :::|| |||  :|||||||
884 CAAAATACTCCGCGTTGCGAATAACCATGCTCTGAGAACCATCAACACTT 933

134 SerThrLysPheGluAspTyrAla.................. 141
    |||||||                 |||
934 ATCTATACTTTCACAGATTATATTGCCGGGTTTAGATAAAGTCCAGTTGTC 983

141 ........................ 141

984 TGCAGAATTGAGCTTTATTCCTAGAGAATAAGGAGTGTTGGAAAATACTA 1033

142 .........MetSerProArgIleThrGlyAspGluLeuAsnGlnLys 154
    :::::     |||:||| ||::::|:|||:||||||||
1034 GTATCTCAAATTTTAAGAGTACCATAGGTGGGCAGGAGGATCACCTATAAA 1083

155 .....LeuArgAlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGl 169
    :::                   ||||::||:||
1084 GGAACGGTTAATGTTCTTTTATGGAAATGAGAGCACTAAAGAAGCAATTA 1133

169 YIleMetGluGlyLeu................GluProLeuAsnA 179
    |||                            |||
1134 TATTACTAATGGAATTGAGCAATGGTGGGTGGGAGTATTGAAAGCTACAACA 1183

179 laIleArgValThrGlnGluAlaValaThrTrpTyrTyrSerAspAsnAlaPro 195
    ::| |||       :|||||||::||||||||||
1184 CC.......GAAACGGGAGAATTGTCTGGTATGTTATATGTCAATCCAAAC 1227

196 IleSerAsn....ProAspGluSerPheLys.................. 204
    :::||| |                 |||
1228 CGTACCAAATATTCCTTTACCGACACCATGAATTATTGGGGATTTGGAAGGGC 1277

205 .ArgGluSerGluSerAsnLeuValSerThrSerGlnLeuSerLeuMetA 221
    :::::     |||:| |||||::|:|||||:
1278 TCGTTCAAAATACAACGCGACTTAGAAAAACTTACTTTAAGAAGATATCAC 1327

221 rgGlnAlaLeuLysGlnLeuIleAspProAsnLeuAlaThrLysMetPro 237
    ||:                          |||
1328 AGCTTGGAGAGGATTCAGGTCGTTATGAAGTACCTGAAGGAGAAAAATTACCA 1377

238 LysGlnValProAspAsp....................... 243
    |||                            |||
1378 TCAAGTTATGGGGTTTGATGTTACAAAAACTTACTTTAAGAACGGATATCAC 1427

244 ...........PheGlnLeuSerIlePheGluSerGluAspL 254
    |||||||||                     |||
1428 AGCAGGCCTAGGAAATGGTTTCAAATGACCAAACGTCAGCGAATTGACT 1477

254 ysGlyAspLys....TyrAsnLysGlyTyr.............. 262
    |||||||||                   |||
1478 TTGGAAATAATATCCAAAAATAAACCATTTATCATCAAAGTAACAGGGAAA 1527

263 ...........GlnAsnLeuLeuSer...... 267
    |||                       |||
1528 ACAGACCAATCTGGTAAGCCACATTGGTTGTTCAATCCAATTTGGCAAGTTT 1577

268 ..................GlyGly......... 269
    |||                           |||
1578 TCGTGGTGGTTGCTTCTGAATAATGCTGCTTTTACTCCAGTTGGAGGAAATGTCT 1627

270 .......LeuValProThrLysProProThrProGly 279
    |||||||                        |||
1628 ACTTCCAAAACGAAATTGCCTTGTCTCCTTCTAAGGGTAGTGGTTCTGGG 1677

280 AspProProMetProProAsnGlnProGlnThrThrSerValLeuIleAr 296
    ||| ||||||:||||||||:|::|:|||||||
1678 AAAAGTGAATTT........ACTAACGCCCTCTATTACAGTAGCAAAATCTAAA 1721

296 gLysTyrAlaIleGlyAspTyrSerLysLeu.............. L 307
    |||||||                      |||
1722 ACGAGTGGGCTCAGCTTCGCGTTTAAGAAATTGTCAACTGACAAATGTCCAAT 1771

307 euGluGlyAlaThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGln 323
    ::|||| |:::||| :||||||||:||||||||||
1772 TGCCGAAGGCGGGCTTTTTGAGCTGCGTTCATCAAAAATGGTAATAGT...CAG 1818
```

```
seq_name: /cgn1_8/cggdata/geneseq/geneseqn/NA2000.DAT:A97541

seq_documentation_block:
ID   A97541 standard; cDNA; 2700 BP.
XX
AC   A97541;
XX
DT   29-JAN-2001 (first entry)
XX
DE   Human Hsp72 (heat shock protein 72) cDNA.
XX
KW   Human Hsp72; heat shock protein 72; chromosome 6p21.3; Hsp72 inhibitor;
KW   expression modulator; JNK phosphatase inhibitor; antiproliferative;
KW   drug screening; cancer; leukaemia; lymphoma; solid tumour; sarcoma;
KW   carcinoma; breast cancer; prostate cancer; premalignant condition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200054814-A1.
XX
PD   21-SEP-2000.
```

```
PF
XX
PR   17-MAR-2000; 2000WO-US07350.
XX   18-MAR-1999;    99US-0125046.
PA   (PHYL-) PHYLOGENY INC.
XX
PI   Volloch V2, Sherman M;
XX
DR   WPI;  2000-647056/62.
DR   P-PSDB; B23252.
XX
PT   Identifying compounds that inhibit proliferation of cells and capable
PT   of modulating the expression of heat shock protein 72 gene and/or
PT   activity of Hsp72 useful for treating cancers such as leukemia,
PT   lymphoma -
XX
PS   Examples; Fig 16A; 77pp; English.
XX
CC   The invention relates to a novel method of identifying compounds that
CC   inhibit proliferation of cells comprising contacting a test compound with
CC   a cell which overexpresses Hsp72 (heat shock protein 72), and determining
CC   if the test compound inhibits activity or expression of Hsp72.
CC   Optionally, Hsp72 is contacted with the test compound under optimum
CC   conditions to allow the two components to interact and bind, forming a
CC   complex which is detected. The invention also relates to a method of
CC   identifying compounds that inhibit Hsp72-mediated JNK phosphatase
CC   activation, comprising contacting a test compound with a cell which
CC   expresses Hsp72, exposing the cell to a heat induced stress and
CC   determining if the compound inhibits JNK phosphatase activity. The
CC   invention additionally encompasses compositions comprising an inhibitor
CC   of Hsp72 or JNK phosphatase activity. The compounds identified as
CC   inhibitors of Hsp72 or JNK phosphatase activity are useful for
CC   inhibiting the proliferation of cells. Modulation of the activity of the
CC   JNK phosphatase or Hsp72 is used to treat a proliferative disorder such
CC   as cancers (e.g. leukemia, lymphoma, solid tumours such as sarcomas and
CC   carcinomas, breast cancer, prostate cancer). The compounds that inhibit
CC   Hsp72 activity can also be administered to treat premalignant conditions
CC   and to prevent progression to a neoplastic or malignant state. The
CC   compounds that inhibit Hsp72 function are administered to a patient
CC   having a disease or disorder mediated by an increase of Hsp72 expression
CC   or activity relative to normal levels. The present sequence represents
CC   cDNA encoding human Hsp72 used in the exemplifications of the invention.
XX
SQ   Sequence 2700 BP; 601 A; 780 C; 834 G; 485 T; 0 other;
```

alignment_scores:
```
Quality:          149.00          Length:          655
Ratio:              0.463         Gaps:             34
Percent Similarity: 49.160        Percent Identity: 22.137
```

alignment_block:
US-09-494-297-2 x A97541

```
Align seg 1/1  to: A97541   from: 1  to: 2700

US-09-494-297-2 x A97541

141 AlaMetSerProArgIleThrGlyAspGluLeuAsnGlnLysLeuArgAl 157
    |||       |:||                ::||            |||
189 GCCTCATCGAGCTCGGTGATTGGCTCAGAGGGAAAAAGGCGGGTCTCCGT 238

157 aValMetTyrAsn..........GlyHisProGlnAsnAlaAsnGlyIleMetG 172
    :::|                   |||                  |
239 GACGACTTATAAAACGCCAGGGCAAGCGGGTCGGATAAACGGCTAGGCTG 288

172 luGlyLeuGluProLeuAsnAlaIleLeuArgValThrGlnLeuAlaValTrp 188
    ||        |||             ||||:|
289 AGGAGCTG...............CTGCGACAGTCCACTACCTTT 317

189 TyrTyrSerArgAspAsn........AlaProIleSerAsnProAspGluSerPheTyr 204
    ||||||:||:|||            ||::|     |     |||
318 TTCGAGAGTGACTCCCGTTGTCCCAAGGCTTCCCAGACGCGAACCTGTGCG 367

204 sArgGluSerGluSerAsnLeuValSerThrSerGlnLeuSerLeuMetA 221
    ::       :|
463 euPheLysTyrThrValLysProArgAspThrAspProAspThrPheLeu 479
    |||       :::
```

Right column:
```
368 GCTGCAGGCGACACCGGCGCGTGCGAGTTTCCGGC............... 402
    :|||||                                       :::|
            ::::: ::                  |||||||        GTCC

221 rgGlnAlaLeuLysGlnIleLeuAspProAsnLeuAlaThrLysMetPro 237
    :|||:|  |||||||||  |:|                |||
403 GGAAGGACCGGAGCTCTTCTCGCGGATCCAGTGTTCCGTTCCAGCCCCCA 452

238 LysGlnValProAspAspPheGlnLeuSerIlePheGluSerGluAspIy 254
    |:|              |||||::::||::::                :::
453 ATCTCAGAGCCGGAGCCCGACAGAGACAGGGAACACGGCACAAAGCCG 502

254 sGlyAspLysTyrAsnLysGlyTyrGlnAsnLeuLeuSerGlyLysLeuV 271
    :|||::::|:|::|||||   ||    |||:::       ||| |||
503 CGGCGGATCGGCATCGACCTGGGCACCACCTACTCCTGCTGGGGT...G 549

271 alProThrThrLysProProGlyAspProProMetProProAsnGln 287
    ||||||||||:::                          |||::::|||
550 TTCCAACACGG.......CAAGGTGGAGATCATCGCCAACGACCAGGCAA 593

288 ProGln..ThrThrSerValLeuIleLeuArgLysGlyTyrAlaLeuAspTyrS 304
    |||:||  |||      |||||||||||||||:||||         :
594 CCCGAACCCCCATCCCCAGCTACGTG..........GCCTTCACGGACACG 631

304 erLysLeuLeu.....GluGlyAlaThrLeuGlnLeuThrGlyTyrAspAsnVal 319
    :::|||:::|||      |||||||:||||      |||||:|    :|
632 AGCGGCTCATCGGGATCGGGCCCAAGAACCACGGTGGCGTGACCCGCAG 681

320 AsnSer......PheGlnAlaArgValPheSerSerAsnAspIleGlyGlur 335
    |||:::     ...|||:::|||:|||||   :::|||:::|||:::
682 AACACCGTGTTGACGCGAAGCGCCTGATTGGCCGCAAGTTCGGCGACCC 731

335 gIleGluLeuSerAspGlyThrThr.......ThrLeuThrGluL 348
    |  |||||                      :::|||||||||
732 GGTGGTGCAGTCGGACAGAAGCACTGCGCCTTTCCAGTGTGATCAACGACG 781

348 euAsnSerProAlaGlyTyrSerIleAlaGlyLeuProIleThrPheLysVal 364
    :::|||||||||||||||||||||||     |||||
782 GAGACAAGGCC..............AAGGTGCAGGTGAGCTACAAGGGG 816

365 GluAlaGlyLysValThrTyrThrIleIleAspGlyLysGlnIleGluAsnPr 381
    |||:::||||     |||||||:||::||||:||||:::|||   :::
817 GAGACCAAGGCATTCTACCCGGAGGAGATCTCGTCCATGGTGCTGACCAA 866

381 oAsnLysGluIleValGlyProGlyTyr.................. 389
    |||||||||:::||:::||:||||     |||
867 GATGAAGAGGATCGCCGAGGCGGTACCTGGGCTACCCGGTGACCAACGCGG 916

390 .....SerValGluAlaTyr....AsnAspPheGluGlu.......... 399
    :::|||  |||||||||  ||||||                  |||||
917 TGATCACCGCTGCCGGCCCTACTTCAACGACTCGGCCAGCCGCCACCAAG 966

400 ...............................PheSerValLeuThrThrGl 406
    :::                             ::|
967 GATCGCGGGTGCTGATCGCCGGGGGCTCAACGTGCCTCGGCGATCATCAACGAGCC 1016

406 nAsnTyrAlaLysPheTyrTyrAlaLysAsnLysSerSerGln. 422
    |||:::|||||||||||||||:::::|::::||||::|||
1017 CACGGCCCGCGCCCATCGCCTACGGCCTTGGACACAACGGGCAGGGGAGC 1066

423 .ValValTyrCysPheAsnAlaAspLeuLysSerProAspSerGlu 438
    :::|||  |||:::      |||||||
1017 GCAACGTGCTCATCTTT.....GACCTG................ 1089

439 AspGlyGlyLysThr.................MetThrProAspPheThrTh 450
    |||||||||||     :::                  |||
1067 GCAACGTGCTCATCTTT.....GACCTG................ 1089

450 rGlyGlyValLysTyr...........ThrHisIleAlaGlyArgAspL 463
    |||||||||||||       ::||||:::||:|:|||||:|||
1090 ...GGCGGGGGCAACCTTCGACGTGTCCATCCTGACGGATGACGACGGCAT 1136
1137 CTTCGAGGTGAAGGCCCACGGCGCCGGGACACCCACCTGGGTGGGGGGACT 1186
```

```
seq_name:  /cgnl_8/gcgdata/geneseqn/NA1999.DAT:X77593
seq_documentation_block:
```

```
1187 TTGACAACAGGCTGGTG.........................AACCACTTCGTG 1215

 480 LysHisIleLysLysValIleGluLysGlyTyrArgGluLysGlyGlnAl 496
      :::|||::::::|||::||        ::|||::
1216 GAGGAGGTTCAAGAAGAAAAACACAAGAAGAAGGACATTCAGCAACAGCGAC 1265

 496 alleGluTyrSerGlyLeuThrGlnLeuArgAlaAlaThrGlnL 513
      ||||::||::::|||||||::::
1266 CGGTGAGG.........CGGCTGCGCACCCGCCTGCGACA 1294

1266 CGTGAGG.........            |||
      LysLeuLysAspTyrHisGlyPhe.........    535

 513 euAlaIleTyrTyrPheThrAspSer.......AlaGluLeuAspLysAsp 527
      ||||:::|||||||:::         |||||||::::|||
1295 GGGCCAAGAGAGGACACCCTGTCGTCCAGCACCCAGGCCACCTGGAGATCGAC 1344

1345 TCCCTGTTTGAGGGCATCGACTTCTACACGTCCATCACCAGGGCGAGGTT 1394
      LysLeuLysAspTyrHisGlyPhe.........           535
                              |||
 528 LysLeuLysAspTyrHisGlyPhe.........                535

 536 .........GlyAspMetAsnAspSerThrLeu..AlaValAlaL 547
              ||||||||:::|::||||        |||||||
1395 CGAGGAGCTGTGTCCTCCAACCTGTTCCAAGCACCCTGGAGCCCGTGAGA 1444

 547 ysIleLeuValGlnTyrAlaGlnAspSerAsnProProGlnLeuThrAsp 563
      ||  |||   :::      :::       |||::: |||
1445 AGGCTCTGCGCGACGCCAAGCTGGACAAG.........GCCCAGATTCACGAC 1488

 564 LeuAspPhePhe.........IleProAsnAsnAsnLysTyrGl 575
      :::||::|                 ||||||
1489 CTGGTCCTGGTCGGGGGGGCTCCACCCGCCATCCCC.........AAGGTCCA 1529

 575 nSerLeuIleGlyThrGlnTrpHisProGluAspLeuValAspIleIleA 592
      ::||:::|||||||||::::|::||||:::|::
1530 GAAGCTGCTGCAGGACTTCTTCAACCGCGGCCACCCTGAACAAGAGCATCA 1579

 592 rgMetGluAspLysLysGluValIleProValThrHisAsnLeuThrLeu 608
      :: |||                     ::|||||
1580 ACCCCGACGAG.........GCTGTGGCCTACGGGCCGGGGGGGGTG 1614

 609 ArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPheHisPh 625
      :::|||       |||||||::::
1615 CAGGCGGCCATC.........CTGATGGGGGGACAGAGTCCGAG. 1647

 625 eGluIleGluLeuLysAsnAsnLysGlnGluLeuSerGlnIleThrValL 642
      :: |||                :::|||||
1648 .........AACGTGCAGGACCTGCTCTGCTGCTGGACGTGG 1678

 642 ysThrAspLysThrAsnLeuGluPheLysAspGlyLysAlaThrIleAsn 658
      :::|:|||||||
1679 CTCCCCTGTTCGCCTGGGAGACGGCCGTGGAGGCGTTGATGACTGCCCTG 1728

 659 LeuLysHisGlyGluSerLeuThrLeuGlnIleGlyLeuProGlyTyrSe 675
      ::|||:::||::|:::|||||
1729 ATCAAGGGCAACTCACCAATCCCCACCAAG.........CAGACGCAGAT 1769

 675 rTyrLeuValLysGluThrAspSerGluGlyTyrLysLysValLysAsnS 692
      :: ::        |||||||:::
1770 CTTCACCACCTACTCCGACAACCACCCGGGTGCTTAGCTCCAGGTGTACG 1819

 692 erGlnGluValAlaAsnAlaThrValSerLysThrGlyIleThrSerAsp 708
      :: ::|||        |||||||:::
1820 AGGGCGAGAGGCC.........ATGACGAAAGAC 1845

 709 GluThrLeu.........AlaPheGluAsnAsnLysGluProValValProTh 723
      ::  |||||        |||||
1846 AACAATCTGTTCGGGGGGCGGCCATCCCTTCCGCGGCCCCAG 1895

 723 rGlyValAspGln 727
      ||||||
1896 GGGGCGGTGCCCAG 1908
```

```
ID    X77593 standard; DNA; 3945 BP.
XX
AC    X77593;
XX
DT    09-AUG-1999  (first entry)
XX
DE    S. aureus SdrD DNA.
XX
KW    Fibrinogen-binding protein; alpha chain; beta chain; ClfB; SdrC; SdrD;
KW    SdrE; fibrinogen; medical device; competitive inhibitor; pharmaceutical;
KW    treatment; infection; septicemia; osteomyelitis; mastitis; endocarditis;
KW    extracellular matrix; vascular graft; vascular stent; vaccine;
KW    intravenous catheter; artificial heart valve; cardiac assist device;
KW    antibacterial; ss.
XX
OS    Staphylococcus aureus.
XX
PN    WO9927109-A2.
XX
PD    03-JUN-1999.
XX
PF    25-NOV-1998;  98WO-US025246.
XX
PR    31-AUG-1998;  98US-0098427.
PR    26-NOV-1997;  97US-0066815.
XX
PA    (EIDH/)  EIDHIN D N.
PA    (FORF-)  FORFAS T/A BIORESEARCH IRELAND.
PA    (FOST/)  FOSTER T J.
PA    (HOOK/)  HOOK M A O.
PA    (INHI-)  INHIBITEX INC.
PA    (JOSE/)  JOSEFSSON E.
PA    (PATT/)  PATTI J M.
PA    (PERK/)  PERKINS S E.
XX
PI    Eidhin DN,  Foster TJ,  Hook MAO,  Josefsson E,  Patti JM;
PI    Perkins SE;
XX
DR    WPI; 1999-357844/30.
DR    P-PSDB; Y08642.
XX
PT    Staphylococcus aureus fibrinogen-binding proteins for treating
PT    septicemia, osteomyelitis, mastitis or endocarditis
XX
PS    Claim 10; Fig 8; 143pp; English.
XX
CC    This invention describes novel Staphylococcus aureus fibrinogen-binding
CC    proteins that bind both the alpha and beta fibrinogen chains. The
CC    proteins (and their encoding nucleic acids are ClfB, SdrC, SdrD and
CC    SdrE). Staphylococcus aureus is thought to utilize fibrinogen to adhere
CC    to medical devices, binding proteins that bind both the alpha and beta
CC    fibrinogen chains (ClfB, SdrC, SdrD and SdrE) can therefore be used as
CC    competitive inhibitors to block this binding. Antibodies against ClfB,
CC    SdrC, SdrD and SdrE inhibit ClfB, SdrC, SdrD and SdrE mediated binding.
CC    The proteins of the invention can be used in a pharmaceutical composition
CC    for the treatment of Staphylococcus aureus infection e.g. septicemia,
CC    osteomyelitis, mastitis or endocarditis or to inhibit the binding of
CC    S. aureus to the extracellular matrix. The proteins or their fragments
CC    may be used to coat a medical device to reduce the S. aureus infection of
CC    an indwelling medical device, especially where the medical device is
CC    selected from the group consisting of vascular grafts, vascular stents,
CC    intravenous catheters, artificial heart valves, and cardiac assist
CC    devices. ClfB, SdrC, SdrD, SdrE, or an active fragment, subdomain or
CC    encoding gene may be used as a vaccine. The DS (aspartate serine) repeat
CC    region or a gene encoding it may be used as an identifying probe for the
CC    identification of genes and encoding proteins from Staphylococcus aureus
CC    (other than ClfA), S. hemolyticus, S. lugdenensis, and S. schleriferi.
CC    The proteins of the invention have antibacterial activity.
XX
SQ    Sequence 3945 BP; 1617 A; 605 C; 714 G; 1009 T; 0 other;
```

```
alignment_scores:
```

```
                    Quality:  149.00
                    Ratio:    0.389
Percent Similarity: 38.147    Percent Identity: 20.219

                    Length:   1004   Gaps:   45

alignment block:
US-09-494-297-2 x X77593

Align seg 1/1  to: X77593  from: 1  to: 3945
```

```
8   AsnLysLeuAsnThrLeuAsnThrGlnArgValLeuSerLysAsnSer..        23
    |||:|::    ||||     |||::|||      ||||:||||:||||
622 AATTCAAATAATGAAAATGAAAAATGACGAATATCAATTTGCCAAAAGTACAGC   671

24  ...LysArgPheThrValThrLeuValGlyValPheLeuMetIlePheA        39
       |||  ||      ||||::|||||  ||||||   |
672 ACCTAAAACGTTTTGAATACAACAGA.......ATGCGTATAGCAG         706

39  laLeuValThrSerMetValGlyValAlaLysThrValPheGlyLeuValGlu     55
    |||  :::        ||||||:||||:||||::||||:|||
707 CAGTACAGCCCATCATCAACACAGAGCCTAAAAAAGTTAATGATTTAATCACA     756

56  SerSerThrPro.......                                        63
    |||:::::|||                        |||  |||
757 TCAAATACACATTAACTGTCGTGTTGATGCAGATAAAAAACAATAAAATCGT      806

63  nPro.....AspSerSerGluTyrArgTrpTyrGlyTyrGluSerT            78
    |||              ||| :::||||:||||:|||::|||||
807 ACCAGCCCAAGATTATTTATCATTAAAATACAAAATTACAGTTGATGACA        856

78  yrValArgGlyHisProTyrTyrLysGlnPheArgValAlaHisAspLeu        94
    |||  ::: |||        |||||::::|||:|||:::|||
857 AAGTAAAATCAGTTGATTATTTC.....                              879

95  ArgValAsnLeuGluGlyLysSerArgSerTyrGlnValTyrCysPheAsnLe     111
    ::::        |||||||||||:||::||||:|||:::||||||
880 ....ACAATTAAATACTCAGATACAGTACAAGTATAGTAGGATTGAATCC        923

111 uLysLysAlaPheProLeuGlySer.......                         124
    |||:|||:||||            ||| :::||:
924 GGAAGATATTAAAAAATATTGGTGATAATTAAAAGATCCAAATAAAGGTCGAAA   973

124 ysLysTrpTyrTyrLysGlnHisAspGly...........                 136
    |||                      ||| |||:
974 CAATTCGGACTGCAAAACATGTACTCCAAAATAAATTTAATTACATATACA     1023

137 PheGluAspThrAlaMetSerProArgIleThrGlyAspGluLeuAsnGl        153
    |||     |||||:|||||::|||        |||||::|||
1024 TTTACAGATTATGTT.......                                  1052

153 nLysLeuArgAlaValMetTyrTyrAsnGlyHis.......                164
    |||||:||||||           ||||::|||||:    P
1053 TGTACAAAATGGGAATTTAATTCAATTTATATGGATGCTGATACAAATTC      1102

164 roGlnAsnAlaAsnGlyIleMetGluGlyLeuGluProLeuAsnAlaIle        180
    |||        |||:|||        :|| ||:|
1103 CTGTTAGTAAAAACGANGTTGAGTTTAAGTTACGATAGGTAATACTACA       1152

181 ArgValThrGlnGluAlaValTrpThrTyrTyrSerAspAsnAlaProIleSe     197
    |||    ::: ||||||:|| |||:|||||:
1153 ACAAAAACAACTGCTAACATTCAATATCCAGATTAATGTTGTAAATGAGAA      1202

197 rAsnProAspGluSerPheLysArgGluSerGluSerAsnLeu.......        211
    |||        |||       ||| |||:
1203 AAATTCAAATTGGATCAGGTTCACTCGAAACAGTTTCACATGTTGGAAATA     1252

212 .......ValSerThrSerGln                                    216
       |||:||   |||:|||:||
1253 AAGAAAATTCCAGGGTACTATAAAACAAACGATTTTAGTAAATCCATCGGAA    1302

217 LeuSerLeuMetArgGlnAlaLeuLys.......                       229
    ||||||||||||          |||  |||:||:||
1303 AATTCTTTAACAAATGCAAAACTAAAAGTTCAAGCTTACCACTCAAGTTA      1352
```

```
229 pProAsnLeuAlaThrLysMetProLysValProAspAspPheGlnL          246
    ||||||||        |||:||        ;
1353 TCCTAATAATATCGGGGCAAATAAATAAAATGAAAAGATGTAAACAGAT        1393

246 euSerIlePheGluSerGluAspLysSerGlyAlaAspLysTyrAsnGlyTyr     262
    :::||||||||:||:: ||      ||||||||||||||:|::|||    A
1394 TAAAAATATATCAA..GTTCCTAAAAGGTTATACATTAAATAAAGGATAC    1440

263 GlnAsnLeuLeuSerGlyGlyLeuValProThrLysProProThrProGl       279
    ||||||||||| ||||||||:|||:|||||||||||||||||||||
1441 GAT.......GTGAATACTAAAAGGCTTACA.......                  1464

279 yAspProProMetProProAsnGlnProGlnThrThrSerValLeuIleA       296
    |||                      :::|||:::
1465 .......GAATGTAACAAAATCAATACTTGC                         1486

296 rgLysTyrAlaIleGlyAspTyrSerLys.......                     305
    :||| |||    ::: |||||
1487 AGAAAATTACATATGGGCACAACAATACGCGCTGTTATTGATTTTGGAAAT    1536

305 .......                                                   305
1537 GCAGAATTCTGCTTTATGTTGTTAATGGTTAATACAAAAATTCCAATATACAAA 1586

306 LeuLeuGluGlyAlaThrLeuGlnLeuThrGlyA                       317
    |||:|::|| |||:||:|||:|
1587 TAGCGAAAGCCCAACACTTGTTCAAAATGGCTACTTTATCTTCAACAGGTTA    1636

317 sp.......                                                 317
    ::
1637 ATAAATCCGTTTCTACTGGCAAATGCTTTAGGGATTTACTAAATAACCAAAGT  1686

317 .......                                                   317
1687 GGCGGAGCTGGTCAAGGAGTATATAAAAATTGGTAACTACGTATGGGAGAGA    1736

318 AsnValAsnSerPheGln                                        323
    ||| ||||| |||
1737 TACTAATAATAAAACGGTGTTCAGAATTAGGAGAAAAAGGCGTTGGCAATG     1786

324 AlaArgValPheSerSerAsn.......                             335
    |||||||||    |||||||:|
1787 TAACTGTAACTGTATTTGATAATAATAATACAAAATACAAAAGTAGGAAGCA    1836

336 IleGluGluSerAspGlyTyrThrTyrThrLeuThrGluLeu.......        348
    |||:||||||::| |||       |||||:|||:
1837 GTTACTAAAGGAGATGGGGTCATACTTGATTCCAAACTTACCTAATGGAGA     1886

349 AsnSerProAlaGlyTyrSerIleAlaGlu.......                    358
    ||| ||| |||:||||||||:||:|
1887 TTACCGTGTAGAATTTTCAAAACTTACCAAAAAGGTTATGAAGTAACCCCTT    1936

358 .......                                                   358
1937 CAAAACAAGGTAATAACGAAGAATTAGAGTTCAAACGGCTTATCTTCAGTT     1986

358 .......                                                   358
1987 ATTACAGTTAAATGGCAAAGATAATAACTTATCTGCAGACTTAGGTATTTACAA  2036

358 .......                                                   358
2037 ACCTAAATATCAAACTTAGGTGTACTATGTCTGGGAAGATACAAATAAAAATG   2086

359 .......ProIleThrPhe                                       362
       ||| :::|||:::
2087 GTATCCAAGACCAAGATGAAAAAAGGTATATCTGGCGTAACGGTAACATTA     2136

363 LysValGluAlaGlyLysValTyr.......                          375
    ||| ||| |||||||||:||
2137 AAAGATGAAAAACGGTAACGTGTTAAAAACAGTTACAACAGACCGCTGATGG    2186

375 yLys.......                                               385
    |||        :::||:||
2187 .......GlnIleGluAsnProAsnLysGluIle.......
```

```
2187 CAAATATAAATTTACTGATTTAGATAATGGTAATTATAAAGGTTGAATTTA 2236
386  ............ValGluProTyrSerValGluAlaTyrAsnAspPhe 397
        :::::::  |||      |||   :::::
2237 CTACACCAGAAGGCTATACACCGACTACAGTAACACTCGGTAGCGACACATT 2286
398  GluGluPheSer...ValLeuThrThrGlnAsnTyr............ 408
     |||||:|||       ||||||||||||
2287 GAAAAAGACTCTAAATGGTTTAACAACACAGGTGTTATTAATGGTGCTGA 2336
409  ..........AlaLysPheTyrTyrAlaLysAsnLysAsnGly 420
        ||||| |||      :::
2337 TAACATGACATTAGATAGTGGGAATTCTACAAAACACCAAAATATAATTTAG 2386
420  erSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProProAsp 436
     :::||||||||     |||||    :::|||
2387 GTAATTAATGTAATGGGAAGATACAAAATAAAGATGGTAAGCCAG...... 2430
437  SerGluAsp.........GlyGlyLysThrMetThrProAspPheThrTh 450
        |||     |||       |||:::|||   ::
2431 TCAACTGAAAAAGGTATTCAGGCGTAACAGTTACATTGAAAAATGAAA 2480
450  rGlyGluValLysThrGlyTyrThrHisIleLeuIleAlaGlyArgAsp 466
     :::|||||||||  |||||||||  |||
2481 CGGTGAAGTTTTACAAAACAACTAAAAACAGATAAAGATGGTAAATATCAAT 2530
466  yrThrValLysProArgAspThrAspProAspThrPheLeuLysHisIle 482
        :::||||||||   ||    |||
2531 TTACT 2535
483  LysLysValIleGluLysGlyTyrArgGlyLysGlyGlnAlaIleGluTyr 499
        |||     ::::|||
     GGATTAGAAAATGGAACTTATAAAGTTGAATT
499  r.........SerGlyLeuThrGluThrGlnLeuArgAsnAlaThrGlnL 513
        :::|||:::|||  |||  |||||
2568 CGAAACACCATCAGGTTACACACCAACAACAGGTTCAGATGGAACTCATG 2617
513  euAlaIle...TyrTyrPheThrAspSerAlaGluLeuAspLysAspLys 528
     |||       |||||||||||  :::
2618 AAGGTAATAGATTCAAAAGGTACATCAACAACAGGTGTCATTAAAGATAAA 2667
529  ............................................LeuLeu 530
                                                     |||
     LeuLy
530  sAspTyr.......HisGlyPheGlyAspMetAsnA 540
        |||||||      :::   |||
2718 TGACTAYGTATCGGGAAGATACAAATAAAAAACCGGTGTTCAAGATAAAGATG 2767
540  spSerThrLeuAlaValAla.................................Lys 547
        ::::||||||                                      |||
2768 AAAAAGGGCATTTCAGGTGTACAGTTACGTTAAAAAGAATGAAAAACGACAA 2817
548  IleLeuVal GlThrTyrAlaGlnAspSerAsnProPro...GlnLeuThrAs 563
        :::|||  |||||||||  ||  |||:::||||
2818 GTTTTAAAAACAGTTACAACAGATGAAAATGGTAAATATCAATTCACTGA 2867
563  pLeuAspPhePheIleProAsnAsnLysTyrGlnSerLeuIleGlyT 580
     |||||  ::::||||   |||
2868 TTTA.....................AACAATGGAACTTATAAA. 2889
580  hrGlnTrpHisProGluAspLeuValAspIleIleArgMetGluAspLys 596
        |||:::||||  |||:::||||
     GTTGAATTCGAGACACCA 2907
597  LysGluValIlePro.......ValThrHisAsnLeuThrLeuArgLys.. 610
     |||      |||      ::::|||
2908 TCAGGTTAYACCAACTCAGTAACTTCTGGAAATGATACTGAAAAAGA 2957
611  ...............ThrValThrGlyLeu....AlaGlyAlaAspArgT 620
        |||||||||      |||  :::
```

```
2958 TTCTAATGGTTTAACAACAACAGGTGTGTCATTAAAGATGGCAGATAACATGA 3007
620  hrLysAspPheHisPheGluIleGluLeuLeuLysAsnLysGlnGluLeu 636
        ||  ||||  ||     |||   |||||  :::
3008 CATTAGACAGTGGTTTC.........TATAAAACACCAAAATATAGTTTA 3048
637  ..............LeuSerGlnThrValLysThrAspLysTh 646
        :::         |||  |||::||||
3049 GGTGATTAGTTTGGTACGACACAGTAATAAAGACGGCAAACAAGATTCAAC 3098
646  rAsnLeuGluPheLysAspGlyLysAlaThrIleAsnLeuLysHisGly 663
        ||   ||||||||  ::|||||||   ||    |||
3099 TGAAAAAGGTATCAAAGATGTTAAAGTTACTTTATTAAATGAAAAAGGCG 3148
663  luSerLeuThrLeuGlnGlyLeuProGlyGlyTyrSerTyrLeuValLys 679
        ||  ::|||  |||||||  |||:::||||
3149 AAGTAAATTGGAACAACTAAAAACAGAAAAATGGTAAAATACTGCTTTGAT 3198
679  ...LeuSerTyrTyrLeuValLysValAsnSerGlnGluValAl 696
        :::|||      |||  :::
3199 AATTTAGATAGCGGGTAAAAATACAAAGTTAATTTATTTTT........ 3242
696  aAsnAlaThrValSerLysThrGlyIleThrSerAspGluThrLeuAlaP 713
        ||:::     ||   |||::|||   |||||
3243 TGGCTTAACAACAAACAGGTACAAATTACAACTGAAGATGAGAAGATGCAG 3292
713  heGluAsnAsnLysGluProValValProThrGlyValAspGlnLysIle 729
        :::|||      |||  :::
3293 ATGGTGGCGGAAGTTGACGTAACAATTACGGATCATGATGATTTCACACTT 3342
730  ...AsnGlyTyr 732
        |||||||||||
3343 GATAATGGCTAC 3354
```

seq_name: /cgn1_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20228

seq_documentation_block:

CC   The present sequence represents a gene isolated from
CC   Enterococcus faecalis. The present invention describes genes, proteins
CC   and antigenic polypeptides isolated from E. faecalis. The proteins can
CC   be used in vaccines for preventing or attenuating an infection caused
CC   by a member of the Enterococcus genus in an animal. They can also be
CC   used for detecting Enterococcus antibodies in a sample. The nucleotide
CC   sequences can be used for detecting Enterococcus nucleic acids.
CC   Products from the present invention can also be used for screening
CC   compounds to identify agonists and antagonists of E. faecalis protein
CC   activity.
XX
SQ   Sequence 6168 BP; 2097 A; 1176 C; 1389 G; 1506 T; 0 other;

alignment_scores:
        Quality: 149.00   Length: 935
        Ratio: 0.369   Gaps: 44
        Percent Similarity: 43.209 Percent Identity: 19.037

alignment_block:
US-09-494-297-2 x X20228

Align seg 1/1   to: X20228   from: 1   to 6168

```
   19 LeuSerLysAsnSerLysArgPheThrValThrLeuValGlyValPheLe   35
      |||::::||||::::|||||:::||:
 3361 CTGAAGAAAACAGCCATGGATGAGACGACTAGTTTAGGAGTTCCCCCATTT 3410

   35 uMetIlePheAlaLeuValThrSerMetVal..............        45
      ::||:||:::||||
 3411 CCAAATTTGGGACCAAGCTAAAACACAAGTCCTACGTGACAGTAG        3460

   46 ..........GlyAlaLysThrValPheGly...............        52
      :|||:::|||
 3461 ATGCCACCGGGGTTATCACATTGGGTGGTTGCCCAAGGGCAATACATT     3510

   53 LeuValGluSerSerThrProAsnAlaIleAsnProAspSerSerSerGl   69
      ||:||||::||:::||||:||||::
 3511 TTGGGTGGAGACAAAAGCACCA.........................     3531

   69 uThrTyrArgThrPThrTyrGlyTyrGluSerTyrValArgGlyHisProTyrTyrTyrL 86
      |:|:||||||::::|||||
 3532 ..GAAGGCTATACACAGTTTCGGACGAATTACCTA.............    3562

   86 ysGlnPheArgValAlaIleHisAspLeuArgValAsnLeuGluGlySerArg 102
      ||:::|||:::||||::||
 3563 AAGGCCCGAGTCATTACTATTGATGAAGAAACTTCAGCCGAGAGACAACA   3612

  103 SerTyrGlnValTyrCysPheAsnLeuLysLysAlaPheProLeuGlySe 119
      ::|:||||
 3613 CCA.....ACCATTATTAAAAACGATGTCAATAAAGTATTTTTAGAAAAT 3659

  119 rAspSerSerValLysLysTrp.........TyrLysLysHisAspG    132
      |||||::|||||
 3660 GGATGAGAAGGGTTAAAAAGTTAGTCAATGCTCGCGTTTAAAATTAGAGCATG 3709

  132 lyIleSerThrLysPhe.............GluAspTyrAlaMetSerProArg 145
      :::::|||
 3710 CCGTAACCACGCCGTTTACTCATTGGGAAGAAGAGTTCCCCTTGCGCCGGAT 3759

  146 IleThrGlyAspGluLeuAsnGlnLysLeuArgAlaValMetTyrAsnGl 162
      ::||
 3760 CGAACC........................................... 3765

  162 yHisProGlnAsnAlaAsnGlyIleMetGlu.....GlyLeuGluPro... 176
      |||||||::|
 3766 ..ACGCGAAATGGCCAGTTAGAGGTGGATTAGAGTTTAAAACCAG      3805

  177 .LeuAsnAlaIleArgValThrGlnGluAlaValThrProTyrTyrSerAsp 192
      |||       :||
 3806 GGCTTTATCAGTTCACAGAAAATCGAAGCACCGAGCAGGCTATCTTTTAGAC 3855

  193 AsnAlaPro.............................            195
```

```
 3856 ACGACCCCCAAACGAGATTCATCGTTGACACACAAAATACGACGGACAAATTCG 3905
      :::::|||
  196 .........................IleSerArgAsnProAspGluSerPheI  204
      :::::|||  ::|
 3906 TGATGTTCATGTCAAAAATGCTTAATTACCAAGGTTCTTGCTGAACTAATTA 3955

  204 ysArgArgGluSerGluAsnLeuValSerThrSerGlnLeuSerLeuMet  220
      ||:::|||  ::::::|||
 3956 AAAAAAGACCCAAGCAGCGAATCCAATTAGCAGGGTGCTGAATTTTCAGTCCTT 4005

  221 .........ArgGlnAlaLeuLysGln.........LeuIleAspProAsnLe  232
      |:||||::||:::|||
 4006 GACACCCACAGGACAAGCAGTTCGAGAACCACTTAGTTTCGGATGCAAACGG 4055

  232 uAlaThrLysMetProLysGlnValProAspAspSerPheGlnLeuSerIleP  249
      |:|::|||
 4056 AAAAGTCACAGTGACGGATTTAGCCCAGGAGGTCGTTTCACGATTGCAGC   4098

  249 heGluSerGluAspLysGlyAspLysThrArgAsnLysGlyTyrGlnAsnLeu 265
      |||    ||||
 4099 .........GTGGAAACCAAAACGCCA.....................     4116

  266 LeuSerGlyGlyLeuValLeuValProThrLysProProThrProGlyAspProPr 282
      ||||:::|||  ||:::|||
 4117 .GCAGGGTACCTTTAAACACTGAACCAAGTGCTTTCACGAATTGCAC      4163

  282 oMetProProAsnGlnProGlnThr.........................  290
      |||
 4164 AAGGGATCGGGGCAAACCAGCAACACAGTTATAGCCAACGGCTAACTTTGTTA 4213

  291 .........ThrSerValLeuIleArgLysTyrAlaIleGlyAspTyr    303
      |:||||:::|||
 4214 ACTATCAAGGCACGGCGTAAATTAATCAAAAAAGATGTGAATGGA        4257

  304 SerLysLeuLeuGluLeuValAlaThrLeuGlnLeuThrGlyAspAsnValAs 320
      |:::||||
 4258 ..CACTTATTAGTGTGGTGCCGACATTTAAAGTGCTTGATGCGAAGGGACA 4304

  320 nSerPheGlnAlaArgValPheSerSerAsnAspIleGlyGluArgIle.  336
      |::|||  ::||
 4305 AACGGATTCAAAACAGGCTTG.......ACGACAAATAATCAAGGGAAAATTGTTG 4351

  337 .........GluLeuSerAspGlyTyrThrTyrThrLeuThrLeuGluLeuAsnSerPro 351
      :|||
 4352 CAGAGCACTTAGCCCCAGAGAAAATATCGGCTTTGTTGAAGAACCAAAGCGCCA 4401

  352 AlaGlyTyrTyrSerIle...AlaGluProIleThrPheLysVal....  364
      ::||::||
 4402 ACAGGCTATTTATTTAAAATACCACGCCAGTCCCAATTTGAAAATTGCTGAGAA 4451

  365 .GluAlaGlyLys....................                368
      ::|||||||
 4452 AAATGCTGGTAAACCAGCGGTCGTAGTGACAACTTGTGAGTT           4501

  369 .........ValTyrThrIleIleAspGlyGlyLysGlnIleGluAsnProAsn 382
      :||||
 4502 ACAAAAGGGGCTTTCCAAAATCGTGAAAAACGAGATAGCCGACCAACCATTA 4551

  383 LysGluIleValGluProTyrSerValGluAlaTyrArgAsnAspPheGluGl 399
      ::||||
 4552 GCAGGTGCTGTT.....TTTGAATTTATAT...GATCACAATAA         4586

  399 uPheSerValLeuThrThrGlnAsnTyrAlaLys..............   410
      ||:|||  :::|||
 4587 ACAATCATTAGGGATTACAGCAACGAGTGGCAAAGATGGCAAAAATATCT  4636

  411 .........PheTyrTyrAlaLysAsnLys................     417
      :::||||
 4637 TTAGAGACTTGGCGCCCAGGTACCTATTATTATTACAAAGAAATCAAAAGCACCA 4686

  418 .........AsnGlySerSerGlnValValTyrCys..............  426
```

```
4687 AAATTACCAGATGGCCCAGATTATATTATTTATTCCTGAATTAGTAAAAGT 4736
427  ....PheAsnAlaAspLeuLysSerProProSerGluAspGlyGlyL      442
          ::::||||||||:|||::||:.:||||||||::
4737 AGAAATTCGTGGTGATTTCAAAAGGTGATCCGGAGATT              4773
442  ysThrMetThrProAspPheThrThrGlyGluValLysTyrThrHisIle 458
          |||                                   ::::
4774 ..........                        TTCCAATTAGGG      4800
459  AlaGlyArgAspLeuPheLysTyrThrValLysProArgAspThrAspPr 475
          ||||||||:||:::||||||||                    ||
4801 AAAGGACGCGCCGTCTTTAAGAAAATTGATCCGAACCCACTTCC       4850
475  oAspThrPheLeuLysHisIleLysLysSerGlyLysTyrArg.       491
          |                    |||::::|:|:|||
4851 AGGAACCGATTTTTTAAATTGTATCGA                        4894
492  ................          GluLysGlyGlnAlaIle        497
4895 TCTTTGAAAGAGAAGTAAACTGCTGAAAAAGATGGTTCATTGGCTATGGAG 4944
498  ................      GluTyrSerGlyLeuThrGluThrGlnLeu 507
4945 GATTTAGGTGCTGGTAGCTATGAATTAGATGAACTGGATGCAACGGATGG  4994
507  uArgAlaAlaThrGlnLeuAlaIleLeuTyrTyrPhe.............  519
4995 CTTATATCGTCAATAAACAAACCATTTATTTTGTAGTGAAGAGAATTCAA  5044
519  hrAspSerAlaGluLeuAspLysLysAspLysLysAspTyrHisGly... 534
5045 ATGATAAACAACCACTAGATGAGTTAGAGTTTGTAAATTATCAAGCAGAA  5094
535  ..PheGly.......AspMetAsnAspSerThrLeuAlaValAlaLy     547
5095 GTAATGGGACGTAAAGTCAACGAGCAAGGTCAAACCTTAGCGGGTGCAGT  5144
547  sIleLeuValGluTyrAlaGlnAspSerAsnProGlnLeuThrAspL     564
5145 TTTTGCAATTTACAATGCCGATGAGCAGAATCAGCCCAAGGTTCACGGA   5194
564  euAspPhePheIleProAsnAsnAsnLysTyrGlnSer.............  576
5195 TAACATTCTTGAATCGTGCAGGAGAAAAAGTTTCTGAAAATAACAACGGAT 5244
576  ................                                    576
5245 AAGACTGGCGAAATTTACGCTAAAAGGGGCATTACGTTTT           5294
577  ................        LeuIleGlyThrGlnTrpHisP       584
5295 AGTGGAAACGAAAGCACCAACAGGCTATCTGTTAGACACAAGCCTACATC  5344
584  roGluAspLeuValAspIleIleArgMetGluAspLysLysGluValIle 600
5345 CA......TTTGATGTAACCGCCCAATTAGGAAAAGAGCAGCCAATT      5385
601  ProValThrThrHisAsnLeuThrArgLysThrValThrGlyLeuAlaGl 617
5386 GCTTTAGGCGGATCTTATCAATTATCAAGGAACTGCT              5421
617  yAspArgThrLysAspPheHisPheGluIleLeuLysAsnAsnLysG    634
5422 ..........CAATTAACCAAAGAAAACGAAACAG                5446
634  lnGluLeuLeuSerGlnThrHisVal.                         648
5447 GTGAAGCATTGCCAGGTGCGGGTGCTTTAAGGTCATTGATGAAACAGGG   5493
649  GluPheLysAspGlyLysAlaThrIle...AsnLeuLysHisGlyGluSe  664
5494 CAAACCGTAGATGGAGACAAACCAATCTGATTGTCTGACAAGCAAGGCAAAGT 5543
```

```
664  rLeuThrLeuGlnGlyLeuProGluGlyTyr.............        674
          |||                        |||
5544 CATTGCGAAAAACTTAGCACCGGGAAACGTATCGTTTTGTGGGAGACACAAG 5593
675  ........SerTyrLeuValLysGluThrAspSerGluGlyTyrLys...  687
          |||||||||::::|||||||...              :::::
5594 CGCCAACTAGCTATCTGTTCTTAATGAAACGCCAAGCCAAGCTTTACGGATT 5643
687  ................                                    687
5644 GCCAAAGACAACCAAGGCAAACCAGCCACTGTGGTTACCTTAAAGACCTTT 5693
688  ................        ValLysValAsnSerGlnGluV      695
          |||||:::||||||:||||:
5694 TATTAAATTACCAAGGTGCTGCCAAGCTGGGTGAAAATTGATCAGCAAAAGA 5743
695  alAlaAsnAla................                          698
          ||| |||
5744 ATGCCTTAGCGAGGTGCTGAATTAAAAGTGACAGATGCAGAGACAGGGCAA 5793
699  ThrValSerLysThrGlyIleThrSerAspGluThrLeuAlaPheGluAs 715
          |||||||||::|||||:
5794 ACTGTCGCTCGTTCATTACGTTCTGACAACCAAGGGTTACTTCAAGTGAA  5843
715  nAsnLysGluProValValProThrGlyValValAspGlnLysIle...AsnG 731
          |||  |||:  |||   :::|
5844 CCACTTACAACCAGGAGAAAATATACCTTTGTGGGAAACAAAAGCACCGGATG 5893
731  lyTyr 732
          |||||
5894 GTTAC 5898
```